

## Durham Research Online

---

### Deposited in DRO:

11 May 2017

### Version of attached file:

Accepted Version

### Peer-review status of attached file:

Peer-reviewed

### Citation for published item:

Kamei, K. and Putterman, L. (2018) 'Reputation transmission without benefit to the reporter : a behavioral underpinning of markets in experimental focus.', *Economic inquiry*, 56 (1). pp. 158-172.

### Further information on publisher's website:

<https://doi.org/10.1111/ecin.12477>

### Publisher's copyright statement:

This is the accepted version of the following article: Kamei, K. Putterman, L. (2018). Reputation Transmission without Benefit to the Reporter: a Behavioral Underpinning of Markets in Experimental Focus. *Economic Inquiry*, 56(1): 158-172, which has been published in final form at <https://doi.org/10.1111/ecin.12477>. This article may be used for non-commercial purposes in accordance With Wiley Terms and Conditions for self-archiving.

### Additional information:

## Use policy

---

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.

Reputation Transmission without Benefit to the Reporter:  
a Behavioral Underpinning of Markets in Experimental Focus

Kenju Kamei and Louis Putterman<sup>\*</sup>

**Abstract:**

Reputation is a commonly cited check on opportunism, but it is often unclear what motivates an agent to report another's behavior when it is easy for the aggrieved individual to move on. In a sharply focused laboratory experiment, we find that many cooperators pay to report a defecting partner without the possibility of pecuniary benefit when this has the potential to deprive the latter of future gains and to help his next partner. We illustrate how a social preference can explain such costly reporting, and also discuss evidence for a role of emotions.

*JEL classification:* C91, D03, D63

*Keywords:* costly reporting, experiment, reputation, prisoners' dilemma, social preference, inequity aversion

---

<sup>\*</sup> Department of Economics and Finance, Durham University (email: kenju.kamei@durham.ac.uk, kenju.kamei@gmail.com) and Department of Economics, Brown University (email: Louis\_Putterman@brown.edu), respectively. Funding for this project came from the Murata Science Foundation, with additional support from Brown University. We wish to thank Cheewadhanaraks Matanaporn for programming of the experiment software, and Benjamin Chowdhury, Kathleen Hu and especially Hui Wen Ng for assistance with the literature review, theoretical analysis and data analysis. We also thank Yan Chen for her hospitality while the authors conducted the experiments at the University of Michigan, and we thank two anonymous reviewers, Pedro Dal Bó and other participants at the Brown University theory lunch, and participants at the 2015 Social Dilemmas Conference at Brown, and at the ESA World Meeting in Sydney, Australia in July 2015 for helpful comments.

## 1. Introduction

People face many situations marked by potential benefits of cooperating with others and accompanying dangers of being exploited by them. The fact that it is sometimes possible to choose interaction partners based on their reputations for cooperativeness and that one might accordingly have an incentive to invest in such a reputation, is an oft-noted factor capable of incentivizing cooperation. But what motivates actors to transmit the required reputational information?

Both sentiment-focused moral philosophers like Adam Smith (1761) and more recent evolutionary psychologists argue that human nature, which the latter view as having evolved during millennia of small band existence, includes innate dispositions towards cooperation.<sup>1</sup> Built-in inclinations that help us to overcome dilemmas of cooperation may include special interest in others' cheating, emotions of anger supporting willingness to expend resources to punish cheaters, and emotions of guilt and shame that function as self-punishments when indiscretions are not detected.

Closer to our topic of information transmission is gossip, said to account for much of human social interaction in every known culture (Dunbar, 2004). A considerable fraction of gossip consists of reports of others' misdeeds. Feinberg *et al.* (2012) report evidence that “individuals who observe an antisocial act experience negative affect and are compelled to share information about the antisocial actor with a potentially vulnerable person,” and that “individuals possessing more prosocial orientations are the most motivated to engage in such gossip, even at a

---

<sup>1</sup> An accessible book-length treatment is provided by Pinker, 2003. Discussions by economists include Bowles and Gintis, 2011. See also Gintis *et al.* (2005) and Wilson (2012).

personal cost” (p. 1015). Such information appears to have a distinct power to grab attention, and psychological normalcy includes a desire to avoid being the subject of negative gossip.<sup>2</sup>

Being the subject of negative reports may also have direct material consequences, if others hesitate to interact with individuals known for past reneging or defection. In modern economies, individuals can choose where to shop, which plumber to hire, and so forth. But the same mobility that puts potential cooperation partners in competition with each other threatens the reputational mechanism if a cost is entailed in spreading the word about an interaction partner’s defection. If one has already decided to leave the untrustworthy partner behind and there is no incentive to encourage her to change, how will others learn whom to avoid? If all victims care only about own material payoffs, if they cannot profitably exchange the information, and if there is any cost associated with conveying it, then there will be no such reporting.<sup>3</sup> However, there may be non-material motivations for the costly reporting of cheating or opportunistic behavior. First, engaging in negative gossip may be a direct source of satisfaction, paralleling reports that pleasure centers in the brain are activated when experimental subjects punish selfish counterparts in a trust game (de Quervain *et al.*, 2004). Second, tipping others off not to naively cooperate with the miscreant may bring satisfaction not for the act itself but thanks to anticipation of the punishment that this may visit on that actor. The large literature showing cooperative subjects’ willingness to spend money to reduce free riders’ earnings in voluntary contribution experiments (Falk *et al.*, 2005; Gächter and Herrmann, 2009; Chaudhuri, 2011) suggests that such motives are widespread, and raises the possibility that punitive motives

---

<sup>2</sup> As Adam Smith (1761) wrote: “Man naturally ... dreads, not only blame, but blame-worthiness.”

<sup>3</sup> Note that in network theory (see, for example, Jackson and Zenou, 2014), the standard assumption is that while creation of links may itself be costly, information travels costlessly among those actors who are linked together, and that information is passed along by default whenever two agents are linked and the informed agent is indifferent to having transmission occur. We could locate no discussion of behavioral or social preference explanations of willingness to incur costs to transmit information, in this literature.

might also motivate reporting. Third, conceivably the victim feels empathy or obligation towards others who are in danger of being victimized, and may accordingly try to warn them.

Given the importance of reputational mechanisms to solving dilemmas of cooperation and the rapid growth of behavioral economics and social preference research using the techniques of experimental economics, one might expect the costly reporting of partners' behaviors to be the focus of numerous studies. Rather than focusing on the decision to engage in costly reporting, however, the most closely related papers have focused on the design of reputation mechanisms or the effects of information transmission. We address this gap by conducting an experiment that investigates both the willingness to pay to report an interaction partner's uncooperativeness, and subjects' beliefs about how common that willingness is. Our experimental design, which builds on familiar prisoners' dilemma stage games, allows us to isolate the willingness to engage in reporting when a personal monetary payoff is absent, so that only desires to punish the individual, to protect her future interaction partner, or to engage in information sharing for its own sake (gossip) are potential motives for reporting. We demonstrate that reporting is significantly less common when it is costly than when free, but that costly reporting does occur often. We show cooperator-defector reporting to be far more common than cooperator-cooperator, defector-cooperator, or defector-defector reporting. Indeed, a majority of cooperators meeting defectors report them despite its cost and lack of personal benefit, in our study. We identify conditions under which such reporting, and its relative frequency across behavior pairings, is consistent with an illustrative model of social preferences, the inequity aversion model of Fehr and Schmidt (1999).

One condition which must be satisfied to make reporting attractive to an inequity averse agent, in our setting, is belief that a sufficient share of the population chooses cooperation.

Although reporting at positive cost is never rational for strictly selfish agents in our experiment, there exist configurations of beliefs such that the same does not apply to cooperating. We conduct an incentivized elicitation of beliefs about others' cooperation and reporting, and with the resulting data we calculate which subjects could be rationally choosing or rejecting cooperation out of simple payoff maximization, and which choose cooperation (or defection) in error or due to a social preference or emotion. Almost all observed decisions to defect in the experiment are explicable by payoff maximization under own beliefs regarding others' cooperation and reporting probabilities. Rational maximization of own payoff also explains many choices to cooperate, but a substantial number of cooperation choices require alternative explanation. In addition to social preference analysis, we keep in mind that our data on reporting, especially, are compatible with more psychological explanations, and we bring to bear evidence on the role of emotions in a final portion of our discussion.

The rest of the paper is organized as follows. Section 2 briefly summarizes related literature, and Section 3 describes our experimental design. Section 4 provides the theoretical predictions and hypotheses under both monetary payoff maximization and our illustrative social preference theory. Section 5 reports results. Section 6 concludes.

## **2. Literature**

Although many decisions to cooperate are explicable by rational and self-interested agents' believing that others' actions may be non-standard (Kreps, Milgrom, Roberts and Wilson, 1982), decisions to engage in costly reporting are never compatible with material self-interest in our set-up. Social preferences and emotions, as discussed for instance by Camerer and Loewenstein (2003) and Sobel (2005), may be required to explain such reporting, if present.

The literature on the experimental study of costly punishment, beginning with rejections in ultimatum game experiments (Camerer and Thaler, 1995; Camerer, 2003) and continuing into work on public goods games with punishment opportunities (Fehr and Gächter, 2000), includes numerous attempts to explain actions without material benefit to the actor. Experiments find that many subjects incur a cost to punish when there can be no material benefit, and that the threat of punishment can reduce or eliminate incentives to free-ride.<sup>4</sup> One way to understand costly punishment in such games is to see punishers as having an aversion to others free-riding or defecting while they themselves cooperate. For such individuals, imposing an earnings reduction on free riders at a monetary cost to themselves delivers a utility gain that offsets their lowered money earnings. While such punishing can be thought of as resulting from a psychological trait of negative reciprocity (Hoffman, McCabe and Smith, 1998; Bowles and Gintis, 2004) perhaps linked to an emotional state of anger, it might also be rationalized or rendered mathematically tractable by a simpler framework of inequity aversion (Fehr and Schmidt, 1999).

From a structural standpoint, costly reporting and costly punishing represent different stages in a possible sequence of actions. To study punishment conditioned on contributions, Fehr and Gächter (2000) provided their subjects with accurate, cost free information on the contributions of individual team members, but this left unaddressed the question of how such information is obtained and shared in practice. Costly reporting or monitoring is a needed precursor to costly punishment except where effort-free mutual observability is an automatic by-product of working together. The motives underlying reporting need not be identical to those for punishment, so the decision deserves study in its own right.

---

<sup>4</sup> For example, Ertan *et al.* (2009), find that *ex post*, individual subjects earn more the more they contribute to the public good when opportunities to engage in costly punishment are available. Reviews of the literature on punishment in voluntary contribution experiments include Gächter and Herrmann (2009) and Chaudhuri (2011).

Several scholars have researched costly reporting in the form of online product reviews (Dellarocas, 2003). Resnick and Zeckhauser (2002) look at the relationship between eBay reviews and sales, as well as the prevalence of and motivation behind reviewing. They find reviewing to be frequent and suggest that the giving of feedback despite the absence of private material gain might be understood as the carrying out of a “quasi-civic duty” or as part of a “high courtesy equilibrium.” Gregg and Scott (2006) find that eBay reviews are a major deterrent to fraud, helping to reduce asymmetry of information between buyers and sellers. Wang (2010) addresses the motivations behind leaving a review, specifically with respect to Yelp, a for-profit business review site. He finds strong evidence that social image and reviewer productivity are correlated. Bolton, Greiner and Ockenfels (2013) study structural issues related to online reviews, conducting an experiment that provides evidence that negative reviews in environments like eBay can be deterred by the fear of retaliation.

But while the boom in e-commerce provides much of the impetus for recent studies of reputation formation, traditional economic interactions not amenable to the same fixes also justify attention to costly reporting. As Abraham, Grimm, Neeß, and Seebauer (2016) write, “there now exists a huge literature that aims at evaluating and designing reputation mechanisms for internet trading” but still “we know surprisingly little about the determinants which allow for the formation of reputation in economic systems.” Like those authors, our main interest is in “traditional economic transactions, i.e. in environments where no mechanisms exist that make reputational information publicly available.”

Abraham *et al.* is one of a few recent papers, including also Gërxhani, Brandts and Schram (2013) and Fehr and Sutter (2016), in which costly reporting without clear private benefit plays a part in an experiment. Gërxhani, Brandts and Schram (2013) study transmission



of information about employee trustworthiness among employers, in one treatment making such transmissions anonymous so that direct reciprocity is ruled out as an incentive. However, their players interact in the same condition for twenty periods, which can give rise to incentives for “reputation building.”<sup>5</sup> Also, their reporting cost is small, partly a by-product of their experiment’s multi-period structure, and they do not focus on motivations to report, including possibly asymmetric motivations to report “bad actors.” Abraham *et al.* (2016) compare dyadic private information transmission at both zero and positive cost within small populations playing a series of 36 trust games. The repeated setting here too contributes to the cost of any single reporting action being quite small, and it creates incentives to build cooperation in reporting, although reporting is much less common in their experiment when costly than when free. Our paper is the first of which we are aware that studies costly reporting in a controlled setting where the reporting cost can be appreciable and private material gain is fully ruled out.<sup>6</sup>

### 3. Experimental Design

Our experiment consists of four main treatments with opportunities to report a counterpart’s decision, in three of which reporting is costly. (An additional costly reporting treatment conducted by strategy method is discussed later.) In each treatment, subjects play two one-shot prisoners’ dilemma games, each with different, anonymous, randomly selected participants. The payoff structures of the first and second games are identical and of equal money

---

<sup>5</sup> If players believe it to be widely believed that some people are “conditionally cooperative,” then they may find it privately beneficial to invest in promoting such beliefs by behaving cooperatively, since this may support a “cooperative culture” within the group until “end game” behaviors set in. See, for example, Palfrey and Prisbrey (1997), Healy (2007), and Kamei and Putterman (forthcoming).

<sup>6</sup> Another study of reporting that effectively rules out private material gain from reporting is Fehr and Sutter (2016). Like us, they explicitly relate their findings to the anthropologically observed human propensity to gossip. Rather than fix the cost, they study willingness to pay to pass on a report, and find considerable willingness to spend amounts averaging about 0.3 Euro, or about 18% of what the third party earns in a period if not sending a message. But unlike in our experiment their costly reporters are third parties rather than parties to the exchanges. Whereas the third party feature is what accounts for absence of material benefit to the reporter, in Fehr and Sutter, it is limited repetition and perfect stranger matching that has this effect in our design.

value to end-of-session earnings. The payoff table of each round in U.S. dollars is summarized in FIGURE 1.<sup>7</sup>

A key feature of our design is that subjects decide whether to play the cooperate (denoted X) or the defect (denoted Y) option in *both* games at the outset. Thus, in the instructions read by the participants, we write XX (YY) to represent the cooperate (defect) option, duplicating the letter choice to indicate its play in two games. In what follows (but not in the subject instructions, which avoid such terms) we refer to a subject who chooses XX (YY) as a cooperator (defector), or occasionally XX- (YY-) chooser. Committing subjects at the outset to a single choice captures the notion that people have tendencies that they carry from interaction to interaction. Adoption of this feature simplifies both analysis and reporting decisions, since it means that reporting, e.g., a defector, can be a reliable warning about the kind of agent the next partner will encounter. Of course, we need to take into consideration that imposition of this rule affects players' strategic calculations. Accordingly, pre-commitment for two rounds of play should not be misconstrued as being a mechanism to force type revelation in the sense of the theoretical literature. Indeed, we will show shortly that under some beliefs, it becomes rational for a strictly self-interested agent to select XX (cooperate).

After being randomly matched with a counterpart, selecting between the two options, and being informed of the outcome of their first round interaction, each player in the reporting treatments decides whether to report the decision of her first counterpart. If a player is reported,

---

<sup>7</sup> Payoffs were quoted in dollars; no “lab currency” was used. While predictions for the PD game are the same over a wide range of payoff configurations, the degree of “temptation” to defect, “fear” of being defected on, and the potential to gain from mutual cooperation relative to mutual defection, are impacted by the specific payoff structure, so that conclusions from an experiment with one payoff configuration may not extend, behaviorally, to alternative payoffs. See, for instance, Ahn *et. al* (2001), and Charness *et al.* (2016). We made both the “temptation” and the “fear” small because we anticipated that predictions about the share of cooperators, and actual inclinations to cooperate, would thus be larger, in turn increasing reporting of defectors.

then the second-period counterpart of the reported player is told what that player chose in the first round<sup>8</sup> and is given the option to change his initial choice of X or Y taking into account the report he received—this being the sole exception to the rule that an initial choice is binding for two rounds of play. Subjects know that they will certainly not play the game a second time with the same counterpart, so reporting in the hope that one might oneself be the beneficiary of the information is ruled out. The counterpart will also be told whether she was reported on by her initial partner, which determines whether the player whose initial choice was reported to her is in a position to select a new action.<sup>9</sup> For example, consider two players, A and B. Suppose that both A and B are paired with other randomly assigned partners (not each other) in the first game. Suppose also that A chooses XX (cooperate) and B chooses YY (defect). Now imagine that B's first round partner decides to report her (B's) decision to B's second interaction partner, namely A. Since B was reported, A has a chance to change his initial choice from X to Y so as to avoid being exploited by B. A is also free to stick to the choice of X. Suppose, finally, that A (the cooperator) is not reported by his initial partner. A is thus informed that B has no opportunity to change her choice, so A knows with certainty that B is playing Y in his interaction with her. In this example, A knows that switching to Y will protect him with certainty and that the choice would carry no danger of foregoing a mutual cooperation payoff.

As another example, consider a cooperator (XX-chooser) C who learns that her second counterpart D had selected XX and has no opportunity to change his decision (C's initial counterpart did not report C's choice). C may wish to switch to Y to exploit D, but might decide

---

<sup>8</sup> That is, the computer delivers a truthful report. The potential issues of deciding whether to report truthfully and whether to believe a report that has been received are thus eliminated as concerns. We discuss the impact of this simplifying element in the conclusion.

<sup>9</sup> To preserve maximum anonymity among the subjects in the experiment, those who had no opportunity to change their own choice were asked to answer a trivia question bearing no relation to the experiment, to keep number of computer clicks consistent across all those in the lab.

to stick with X in order to avoid feeling guilty, experiencing disutility from advantageous inequality, etc. In the otherwise similar situation in which D can change his choice, C would have that information and would need to factor in her belief about the likelihood of D changing to Y (which may in turn be influenced by D's belief about the likelihood of C switching).<sup>10</sup> To check sensitivity of reporting to its cost, we vary it across treatments. A reporting cost of \$1, \$0.50, \$0.05, and \$0 is present in the High Cost (HC), Medium Cost (MC), Low Cost (LC) and No Cost (NC) treatments, respectively.<sup>11</sup>

Subjects are also asked, after their own choice of XX or YY, for their beliefs about the percentage of their peers choosing XX, and —after their reporting decisions—for their beliefs about the percentages of defectors and of cooperators who will be reported. The elicited beliefs extend our ability to explore possible motivations behind subjects' decisions. So as not to raise its salience too much, we do not tell subjects about the presence of the belief elicitation tasks before they make the corresponding choices. Subjects are asked for their expectations regarding behaviors of *other* participants only (themselves not included), to avoid hedging. Eliciting beliefs is incentivized by offering a \$1 bonus payment for guesses that are within 5 percentage points of the actual percentage. At the end of their session, subjects are also asked about emotions potentially affecting their reporting decisions: (i) their level of anger toward their initial partners and (ii) their feelings of obligation to help the third party in the second round via reporting.

---

<sup>10</sup> While our design is one of finite repetition, we view the question addressed—that of lack of monetary incentive to engage in costly reporting—as relevant also to a world of indefinitely repeated interactions, since agents in such environments may also periodically need to seek new interaction partners and may avoid dealing again with an individual found to be opportunistic, but have no selfish material motive for incurring a cost to convey the information to others, especially when interacting in a population so large that inducing reciprocal reporting from others is not a plausible motive. Having only two interactions makes practical relatively large stakes in the lab for each interaction, while the fact that the report affects only one future interaction makes the motivational problem more challenging by limiting the punishment that reporting can inflict.

<sup>11</sup> We suspected that a cost of \$1 would suffice to deter many from reporting in the context of an experiment with expected total earnings in the neighborhood of \$15 to \$25, thought we might see more reporting with that cost halved, and we included the much lower cost \$0.05 in case substantial reporting were forthcoming only at a more trivial cost. We included zero cost to observe the inclination to report when cost is not a factor.

#### 4. Theoretical Predictions without and with Social Preferences

Although our main focus is on costly reporting, we discuss here predictions of subject decisions with regard to both the cooperating and reporting decision, since each may be conditioned by beliefs about the other. We begin with the extreme assumption of strictly selfish preferences, rationality, and common knowledge (SRC), then relax the common knowledge assumption to allow for beliefs that some may be otherwise motivated, and finally relax the selfish preference assumption to allow that the decision-maker herself may have a social preference capable of explaining actual costly reporting and some decisions to cooperate contrary to material self-interest. We leave consideration of emotional factors to be discussed when we view the experimental results.

##### *4.1 Common knowledge of rationality and self-interest*

The standard theory predictions in the experiment (with SRC) are straightforward. In the HC, MC and LC treatments, it is never payoff-enhancing to report, since reporting is costly and players are never matched with the same partner twice. Even if there were to be reporting, which subjects could randomly choose to do in the NC treatment, YY would remain the predicted first choice and Y the free second choice of those receiving reports, thus having no impact on how the PD games themselves are played.

##### *Hypotheses with Self-Interest, Rationality and Common Knowledge (H-SRC):*

*In the HC, MC and LC treatments, each subject chooses not to pay to report choices of her first interaction partner, while reporting occurs randomly in the NC treatment. Subjects in all treatments choose YY (defect). A subject having the opportunity to make a free second choice (which, by the above, occurs only in the NC treatment) always selects Y.*

## 4.2 Dropping common knowledge

Even if a subject is strictly self-interested and rational, belief that others might behave pro-socially and/or that others believe such types exist can lead to a different choice over XX vs. YY. Let  $a_i$  be the fraction of subjects that  $i$  believes will cooperate (select XX),  $b_i$  the fraction she believes will report a cooperating counterpart, and  $c_i$  the fraction she believes will report a defecting one ( $0 \leq a_i, b_i, c_i \leq 1$ ). We solve for the conditions under which  $i$  selects XX or YY under assumptions representing two ends of a continuum of beliefs  $i$  might have about free 2<sup>nd</sup> game choices; we call the two beliefs “pessimistic” and “optimistic,” respectively. If pessimistic,  $i$  assumes that an individual free to revise her second choice always selects Y, in line with self-interest, whereas if optimistic,  $i$  assumes that cooperatively-oriented participants (XX-choosers) will stick with X given an opportunity to make a fresh choice, provided that they are informed that they are meeting another subject who chose XX.<sup>12</sup>

As shown in Appendix A.1, under the pessimistic 2<sup>nd</sup> game belief, we obtain:

$$\text{selfish player } i \text{ cooperates (defects) if } 5a_i(c_i - b_i) + c_i > 2 (< 2) \quad (1)$$

By contrast, under the optimistic 2<sup>nd</sup> game belief, as shown in Appendix A.2, we obtain:

$$\text{selfish player } i \text{ cooperates (defects) if } 5a_i c_i + a_i b_i + c_i > 2 (< 2) \quad (1')$$

A higher fraction of subjects chooses to cooperate based on criterion (1') compared with criterion (1).

### **Hypotheses allowing for Belief in Social Preferences among Others (H-SPO):**

---

<sup>12</sup> Clearly the “optimistic” assumption is strong in that it eschews distinctions based on whether the XX-chooser being met with can also change her choice. Its optimism is nonetheless bounded in that YY-choosers free to revise their second choice are still assumed always to stick to Y. Our data show that some (but not all) cooperators stick to cooperation when having the report of meeting a cooperator, but no defector switches to cooperation in this situation.

*If  $i$  is a material payoff maximizer, but believes that others might cooperate and pay to report (cooperators, defectors) with probabilities  $a_i$ ,  $b_i$ , and  $c_i \geq 0$ , then  $i$  will never pay to report her first interaction partner, and will randomly report or not report if reporting is cost free. Subject  $i$  will choose cooperation (defection) if  $5a_i(c_i - b_i) + c_i > 2 (< 2)$ , assuming that  $i$  has “pessimistic” beliefs about cooperators’ free 2<sup>nd</sup> choices, and will choose cooperation (defection) if  $5a_i c_i + a_i b_i + c_i > 2 (< 2)$ , assuming that  $i$  has “optimistic” beliefs about those choices.  $i$  will always choose  $Y$  if able to make a free 2<sup>nd</sup> choice.<sup>13</sup>*

#### 4.3 Social preferences and decisions to report

Unlike cooperating, costly reporting can never be optimal for an agent aiming to maximize own monetary payoff. Explaining its occurrence thus requires error, emotion, or non-traditional elements in the utility function. We illustrate how the latter can work if the potential reporter  $i$  has the inequity-averse preferences proposed by Fehr and Schmidt (1999):

$$u_i(\pi_i | \pi_j) = \pi_i - \alpha_i \cdot \max\{\pi_j - \pi_i, 0\} - \beta_i \cdot \max\{\pi_i - \pi_j, 0\}, \quad (2)$$

where  $\alpha_i \geq \beta_i \geq 0$ , meaning that aversion to inequality that is unfavorable to the decision-maker (reflected in weight  $\alpha_i$ ) is at least as strong as that to favorable inequality (reflected in weight  $\beta_i$ ), and that the decision-maker never values the latter (“aheadness”) for its own sake ( $\beta_i$  takes no negative values). Also assume that  $\beta_i < 1$ . There are four possible situations under which reporting might occur:

Case 1: subject  $i$  cooperates and learns that her counterpart has also cooperated.

Case 2: subject  $i$  cooperates and learns that her counterpart has defected.

Case 3: subject  $i$  defects and learns that her counterpart has cooperated.

---

<sup>13</sup> Note that by writing conditions (1) and (1') with strict inequalities, we assume that subjects whose beliefs render them rationally indifferent between XX and YY will select the latter, if strictly self-interested. This means that when analyzing our experiment results in section 5, we place the few cases in which (1) or (1') hold with exact equality in the set of observations for which an observed choice of XX requires a social preference explanation.

Case 4: subject  $i$  defects and learns that her counterpart has also defected.

For simplicity, we assume that the only other individual,  $j$ , whose payoff  $\pi_j$  affects  $u_i$  if  $\alpha_i$  (and perhaps  $\beta_i$ )  $> 0$ , is the first interaction partner of decision-maker  $i$ , with respect to whom  $i$ 's decision to engage in reporting is made.<sup>14</sup> Then it can be shown (details are in Appendices A.4) that costly reporting will occur

$$\text{in Case 1 if } 6a_i > \rho, (\alpha + \beta)b_i - \beta > 0 \text{ and } (6a_i - \rho) > \rho/[(\alpha + \beta)b_i - \beta] \quad (3a)$$

$$\text{in Case 2 if } (6a_i - \rho) > \rho/\alpha \quad (3b)$$

$$\text{in Case 4 if } 6a_i > \rho, (\alpha + \beta)c_i - \beta > 0 \text{ and } (6a_i - \rho) > \rho/[(\alpha + \beta)c_i - \beta] \quad (3c)$$

where  $\rho$  is the reporting cost.<sup>15</sup> The analysis in the Appendix shows that the conditions for Case 2 and Case 4 hold regardless of whether  $i$  applies the pessimistic or the optimistic assumption about free 2<sup>nd</sup> choices, while the condition for Case 1 applies only when  $i$  makes the pessimistic assumption; if she makes the optimistic assumption instead,  $i$  will never pay to report. As for Case 3, the analysis indicates that a defector  $i$  will never report a cooperating counterpart if  $\beta_i < 1$ , as assumed above. Even if some individuals have unusually high  $\beta$  values, the individuals in question would be unlikely to be defectors (see section 4.4, below), so costly reporting of cooperators by defectors will rarely if ever occur.

Condition (3b) indicates that the more averse to disadvantageous inequality is the cooperator (the higher her  $\alpha_i$ ), the more others she believes to have chosen to cooperate (the

---

<sup>14</sup>  $i$ 's potential concern for  $j$ 's next partner, whom  $i$  may wish to warn or at least inform of  $j$ 's type, is a plausible additional concern that we discuss later. The simple Fehr-Schmidt model also abstracts from explicit considerations of reciprocity (Rabin, 1993; Charness and Rabin (2002) and concern for others' intentions (Falk and Fischbacher, 2006; Dufwenberg and Kirchsteiger, 2004; Cox *et al.*, 2007) as such. Finally, unlike Saito (2013), we assume that the decision-maker cares only about *ex ante* expected differences (see also Brock (2013) and Kamei (forthcoming)).

<sup>15</sup>  $i$  is indifferent between reporting and not reporting if the right hand inequality in each line holds instead with the equals operator.



higher her  $a_i$ ), and the lower is the reporting cost  $\rho$ , the more likely she is to report a defector. Aversion to advantageous inequality,  $\beta$ , plays no role in this decision. In addition, the condition implies that all inequity-averse cooperators report defectors at zero cost (in the NC treatment), as long as  $a_i > 0$ .

These conditions suggest that if distributions of types (that is, of  $\alpha$  and  $\beta$  values) are no different in each case, then the threshold belief  $a_i$  required for costly reporting is lowest for Case 2.<sup>16</sup> In conditions (3a) and (3c), there is also the further implication that costly reporting is conditional on the belief that others do it (a cooperator [defector] is more likely to report another cooperator [defector] if she has a high belief  $b_i$  [ $c_i$ ]). As for relative frequency of reporting in Cases 1 and 4, the two conditions show reporting to be more likely in Case 4 than Case 1 if  $c_i > b_i$ , an intuitively appealing idea that turns out to be strongly supported by our experimental data. Adding to this the finding that reporting is not predicted for any individual making the optimistic assumption about free 2<sup>nd</sup> choices, we arrive at the implication that, assuming sufficient variation of belief  $a_i$  not systematically linked to preference type, costly reporting should be most common in the case of cooperators meeting defectors (Case 2), with the cases of defectors meeting defectors (Case 4) and cooperators meeting cooperators (Case 1) following in that order.

**Hypotheses on Reporting Decisions with Inequity-Averse Social Preferences (H-R-SP):**

*(i) Costly reporting due to inequity averse preferences is most likely in the case of a cooperator meeting a defector, followed by the case of a defector meeting a defector, with reporting of cooperators by cooperators still less common and that of cooperators by defectors rarely if ever occurring; (ii) defectors (cooperators) are more likely to report their defector (cooperator) counterpart the greater the share of others they believe report (the greater is their belief  $c_i$  [ $b_i$ ]);*

---

<sup>16</sup> Notice that  $\alpha - [(\alpha + \beta)b_i - \beta] = (\alpha + \beta)(1 - b_i) \geq 0$ .  $\alpha - [(\alpha + \beta)c_i - \beta] = (\alpha + \beta)(1 - c_i) \geq 0$ .

(iii) the higher is belief  $a_i$ , and the lower is reporting cost  $\rho$ , the larger the share of individuals who will engage in costly reporting.

#### 4.4 Social preferences, beliefs, and decisions to cooperate

Although decisions to cooperate can occur without a social preference, as shown above, having a social preference may cause an agent for whom neither (1) nor (1') holds to nonetheless select X (XX). Using again inequity-averse utility function (2) for purposes of illustration, assuming beliefs  $a_i$ ,  $b_i$  and  $c_i$ , and adopting in turn the pessimistic and optimistic beliefs about free 2<sup>nd</sup> choices, we work out conditions for cooperation in Appendix A.3, and simply summarize here the conclusion:

##### **Hypothesis on Cooperating with Inequity-Averse Social Preferences (H-C-SP):**

*A subject  $i$  who has inequity-averse preferences will be more likely to choose XX the more others she expects to cooperate (the higher her belief  $a_i$ ), the more averse to advantageous inequality she is (higher  $\beta_i$ ), and the less averse to disadvantageous inequality she is (lower  $\alpha_i$ ).*

## **5. Results and Analysis**

### *5.1 Overview of the experiment*

A total of 172 students (152 in the four main reporting treatments, 20 in the strategy method treatment) participated in ten experiment sessions in 2013 and 2014 at the University of Michigan.<sup>17</sup> 58.7% of subjects (101) were female. No subject participated in more than one session, and the sessions lasted about an hour on average, about half of this time being spent on

---

<sup>17</sup> Sessions were conducted by Kamei while he was Assistant Professor at Bowling Green State University. All subjects were recruited from the University of Michigan experimental lab's subject pool using solicitation messages via ORSEE (Online Recruitment System for Economic Experiments). We aimed to recruit 20 participants to each session, but actual numbers varied somewhat due to differing show-up rates, averaging 17.2.

reading of instructions and answering comprehension questions. The experiment was programmed in z-tree (Fischbacher, 2007). All instructions (for the example of the LC treatment, see the Appendix) were neutrally framed, avoiding terms such as “cooperate,” “trust,” etc. Subjects had to answer a number of control questions (see the instructions) to confirm their understanding of the experiment. Communication between subjects was not permitted. Average earnings were \$20.84, including a \$5 participation fee, with a standard deviation of \$4.16.

Figure 2 summarizes key subject behaviors and beliefs, with panels (a) (at left), (b) (upper right) and (c) lower right displaying cooperation decisions, decisions to report defectors, and decisions to report cooperators, respectively. We begin with our main focus, costly reporting. A glance at panel (b) makes clear that there was much costly reporting, but overwhelmingly reporting of defectors by cooperators, as predicted. Focusing on the left panel of Figure 2(b), we see that some 45 to 65% of cooperators who encountered a defector chose to report when costly, and that reporting occurred considerably more often (almost 90% report) at cost zero (NC treatment) than at positive costs (an overall average of 58.6% report in LC, MC and HC). The remainder of this panel and the panel below it, in Figure 2(c), show that there is also some, but considerably less, costly reporting of cooperators by cooperators and of defectors by defectors, and no costly reporting of cooperators by defectors. Overall, an average of 8.0% of subjects in the cooperator-cooperator, defector-cooperator and defector-defector situations choose to report.

**RESULT 1:** *(a) Costly reporting of defectors by cooperators is common (almost 59% report); (b) costly reporting in other cases is significantly less common (overall, 8%); (c) there is significantly less reporting at a positive than at a zero cost. These results are consistent with H-R-SP.*

Turning to cooperation decisions and expectations, panel (a) of Figure 2 shows that regardless of the presence of reporting costs and their size if present, around 50 to 60% of the subjects choose XX in each of the main treatments.<sup>18</sup> The diamonds and triangles in the same panel indicate that cooperators' average expectation regarding the fraction of others who would choose cooperation was significantly higher (around 70%) than that of defectors (around 40%). The difference is highly significant ( $p < 0.001$ , see Appendix Table B.4). With respect to reporting, the right side of panel (b) shows that the average cooperator believed more defectors would be reported than did the average defector, a difference significant for LC subjects and in all costly reporting treatments pooled ( $p < 0.001$ , see Appendix Table B.4).

The displayed frequencies of cooperation and reporting are clearly inconsistent with H-SRC, which, based on the assumption of rational selfish individuals with common knowledge, predicts neither costly reporting nor cooperation.<sup>19</sup> We now discuss in more detail the reporting decisions, considering in section 5.2 their consistency with the illustrative social preference model discussed above, and discussing evidence for a role of emotions in section 5.3.

## *5.2 Predicting reporting decisions with inequity-averse preferences*

In the three treatments in which reporting is costly, 58.6% of cooperators meeting defectors, 16.6% of defectors meeting defectors, 8.8% of cooperators meeting cooperators, and 0% of defectors meeting cooperators, pay the cost to report their counterpart, an ordering of

---

<sup>18</sup> The fractions of cooperators are not significantly different between any two treatments, according to two-sided, two-sample tests of proportions. See Appendix TABLE B.1 panel (C).

<sup>19</sup> We performed binomial probability tests for the conservative null hypothesis that the probability of choosing XX equals 5%, assuming that errors occur with a probability of 5%. This hypothesis was rejected in each treatment. We also performed binomial probability tests for the conservative null hypothesis that the probability that cooperators report the initial choices of their defecting partners equals 5%, assuming that errors occur with a probability of 5%. This hypothesis was also rejected in each treatment. See Appendix Table B.2.

frequencies exactly matching part (i) of H-R-SP.<sup>20</sup> We use again the Fehr-Schmidt model to illustrate the ease or difficulty of predicting reporting choices with a model of social preference. Although we can't identify individual utility function parameters in order to predict which subjects will report their counterparts, as an exercise we estimate the proportion who would be expected to report based on conditions (3a), (3b), (3c) and self-reported beliefs  $a_i$ ,  $b_i$  and  $c_i$  if we assume that each subject has the same likelihood of belonging to each of the four preference types identified and assigned estimated population proportions by Fehr and Schmidt, in precisely those proportions.<sup>21</sup> These calculations imply that about 68.9% of the cooperators who encountered defectors, 13.5% of the cooperators who encountered cooperators, and 14.2% of the defectors who encountered defectors would engage in reporting at the costs obtaining in their treatments given their beliefs and the prevalence of each type. These predicted shares are rather similar to the shares actually reporting, the similar differences between the high share for Case 2 reporting and the low shares for Case 1 and Case 4, in estimate and reality, being especially remarkable.

We can also estimate multivariate regressions to check for patterns consistent with the reporting conditions of inequalities (3a) – (3c). In Table B.8 of the Appendix, we show estimates of simple linear regressions to predict reporting, where the independent variables are the values of the three beliefs and two treatment dummies to control for reporting cost. The regressions for Case 1 and Case 4 show partial consistency with conditions (3a) and (3c) in that the belief variables  $b_i$  and  $c_i$  obtain positive and significant coefficients in their respective estimates. These coefficients suggest a sort of “conditional cooperativeness” with respect to reporting: the higher

---

<sup>20</sup> The percentage of defectors being reported by cooperators is significantly larger than the percentage being reported in the other action pairings according to two-sample z-tests of proportions using pooled data of the three costly reporting treatments (see Panel B.2 of Appendix TABLE B.1).

<sup>21</sup> Specifically, Fehr and Schmidt estimate that about 30% of individuals have  $\alpha = \beta = 0$ , about 30% have  $\alpha = 0.5$ ,  $\beta = 0.25$ , 30% have  $\alpha = 1$ ,  $\beta = 0.6$ , and 10% have  $\alpha = 4$ ,  $\beta = 0.6$ . See also Table 1 in Fehr and Schmidt (2010).

the fraction of others a subject believes report a player who behaves like her counterpart (one who cooperates, in Case 1, one who defects, in Case 4), the more likely that the subject herself pays to report. The estimate for Case 2 suggests a similar sort of conditionality: among cooperators who meet a defector, those believing that a higher share of defectors are reported are themselves more likely to report.<sup>22</sup>

The estimated coefficients on the expected share cooperating ( $a_i$ ) and on treatment dummy variables are insignificant (in one case marginally significant), however, failing to support expectations (based on conditions (3a) – (3c)) that frequency of reporting would be increasing in  $a_i$  and decreasing in  $\rho$ .<sup>23</sup> H-R-SP is supported, with respect to the relationship between cost and reporting frequency, insofar as there is far more reporting at a cost of zero than at a positive cost, in general, and more reporting at zero cost than at each specific positive cost taken individually (see Panels (A) and (B) of Appendix Table B.1).<sup>24</sup> A finer-scale correlation between reporting cost and reporting incidence is also found insofar as subjects' self-reported *expectations* are ones of greater reporting in treatments with lower reporting cost (Appendix Table B.4), but between the different positive costs there is no correlation with actual reporting frequency.<sup>25</sup>

---

<sup>22</sup> There have been numerous behavioral findings of tendency to perform a pro-social or cooperative act conditional on beliefs that others do so, for example Fischbacher and Gächter (2010) for conditional contributing in public goods games and Kamei (2014) for conditional costly punishing in public goods games with punishment opportunities.

<sup>23</sup> Failure of belief  $a_i$  to obtain a significant positive coefficient, whereas  $c_i$  obtains one, in the regression for the Case 2 data, is inconsistent with condition (3b), which implies that  $a_i$  rather than  $c_i$  should be significant. The estimated coefficient on  $a_i$  is insignificant even in specifications that exclude  $b_i$  and  $c_i$  terms.

<sup>24</sup> The fact that reporting is far greater at zero cost than at a money cost as low as \$0.05 might reflect a 'mental accounting' distinction between money and time costs, or a peculiarity of the zero cost as discussed, for example, by Shampanier *et al.* (2007). This suggests that there could be a substantial numbers of individuals willing to provide online reviews with what they treat psychologically as spare time, but who would be deterred if even a small monetary cost were involved.

<sup>25</sup> We note that insensitivity to reporting cost within the range studied by us is consistent with Fehr and Schmidt's assumptions. Specifically, calculations show that with the frequencies of types ( $\alpha_i, \beta_i$  settings, see note 29) assumed by those authors, little sensitivity to reporting cost is predicted within the range of our treatments. For example,

RESULT 2: (i) *The percentages of subjects engaging in costly reporting are well predicted using Fehr and Schmidt's estimates of the prevalence and strength of aversion to disadvantageous inequality and our subjects' self-reported beliefs about frequency of cooperation. Specifically, costly reporting of defectors by cooperators is by far the most frequent case, with costly reporting of cooperators by defectors (predicted to occur rarely if ever) not observed and reporting in the remaining two cases relatively infrequent.* (ii) *Sensitivity of reporting to its cost is supported by significantly more reporting at cost 0 than at costs \$0.05, \$0.50 and \$1.00, and by close correlation between expectations of others' reporting and specific reporting cost.*

### 5.3 Reporting and emotions

Whereas social preference models like that of Fehr and Schmidt assume rational calculation of utility maximizing choices subject to socially interdependent preferences, observed reporting may sometimes have more emotional underpinnings. We turn to evidence that emotions played a part in our subjects' decisions.

Our first source of evidence that emotions helped to trigger some of the observed costly reporting is an additional treatment that resembles the MC treatment in all respects except that rather than having subjects decide whether to report their initial counterpart's choice after learning what that choice was, they are asked to decide in advance whether to send a costly report if the participant with whom they are matched turns out to have selected XX, and likewise whether to send a report if their counterpart turns out to have selected YY.<sup>26</sup> Of the 20 subjects

---

calculations predict that 70% of cooperators will report a defector in both treatments LC and MC, with a drop only to 66.7% predicted to report defectors in HC. Of course, the insensitivity of reporting frequency to cost in our data may be more a quirk of small numbers than a fundamental finding.

<sup>26</sup> These reporting decisions were taken after each subject had made her own choice between XX and YY and had indicated what percentage of others she expected would choose XX. Following the reporting decisions, subjects stated their beliefs about others' reporting decisions under each contingency. To minimize the impact of the

that participated in this strategy method session, 11 selected XX, similar to the 50% share in the MC treatment. Thanks to the set-up, we got decisions from all 11 about whether they would report if meeting a YY-chooser. Only 2 of the 11 (i.e., 18%) chose to pay the required \$0.50 to report their first counterpart if that person chose YY. In the MC treatment, with the same payoffs and cost but a sequential design in which subjects first learned the actual decision of their assigned counterpart, there were 9 who chose XX and learned the counterpart had chosen YY, and 6 of them chose to engage in costly reporting upon learning of their counterpart's action, i.e. 66.7%. A two-sided two-sample z-test of proportions says that the 18% and 66.7% proportions are statistically significantly different from each other ( $p = 0.028$ ). The substantial difference suggests that “hot” emotion (Loewenstein, 2000) may have played a role in reporting in our main treatments.

Other evidence for a role of emotions comes from the survey that all subjects in the main treatments completed following their decisions. Subjects were asked—following all decisions, so as not to contaminate them—to state how pleased or angry they felt about their first counterpart's decision, and how much (if any) sense of obligation they felt to help their first counterpart's next partner. The answers to the anger question indicate that the greatest amount of anger was felt by cooperators who encountered defectors in their first interaction. Self-reported anger is always significantly less for subjects in other pairings—XX meets XX, YY meets XX, and YY meets YY. Panel (a) of Figure 3 shows that among the cooperators who encountered defectors, average self-reported anger towards the counterpart is significantly higher ( $p = .04$ ) for those who chose to report than for those who did not choose to report. Thus, the strength of

---

expectation elicitation processes, subjects were not informed in advance of the fact that expectations were to be elicited, as in our other treatments.



anger could have been a factor prompting decisions to report, although the *ex post* nature of the answers gives reason for caution.<sup>27</sup>

RESULT 3: *Indications that emotions played a part in costly reporting consist of (i) a significantly smaller share of cooperators choosing to report defectors in a strategy method treatment than in the corresponding sequential treatment, and (ii) a significantly higher self-report of anger towards a defector by reporting than by non-reporting cooperators.*

#### *5.4 Obligation, altruism, and normative motives for reporting*

The post-decision exit survey is consistent with presence of another potential motive for a cooperator to report a defecting counterpart: a desire or sense of obligation to help that individual's next partner (say  $k$ ) avoid being exploited. Panel (b) of Figure 3 shows that more than 50% of those actually reporting indicate having felt an obligation to help  $k$ , versus less than 25% of those not reporting.<sup>28</sup> Self-reported feelings of obligation to help are much smaller in the XX-meets-XX, YY-meets-YY and YY-meets-XX cases (see Appendix Figure B.1).

Although a social preference model might also attach value to helping  $k$ , we did not include this motive in our simple illustration using the Fehr-Schmidt model because, with its focus on inequalities and inattention to intentions and other sources of deservingness, that model predicts reporting to help this third party (the counterpart's next partner) only under a narrow range of cases. An alternative explanation, in which reporting to help  $k$  is motivated by belief

---

<sup>27</sup> The anger variable is a subject's response to the question: "How did you feel about your first counterpart's decision? Please rate on a scale from 1 = very pleased to 7 = very angry." That many people view defecting against a cooperating partner as grossly unfair is suggested by the fact that more than half of unaffected *third parties* chose to incur a cost to punish a unilateral defector in a laboratory experiment by Fehr and Fischbacher (2004).

<sup>28</sup> The obligation to help variable is a subject's response to the question: "Did you feel a sense of obligation to help your first counterpart's next counterpart by sending a report? Please rate on a scale from 1 = did not feel obligated at all, to 7 = felt strongly obligated."

that cooperators not only cooperate but report conditional on others doing the same, is discussed in the Appendix and is consistent with Table B.8's finding that for every 1% increase in the proportion of subjects whom  $i$  believed would report a YY-chooser, there is an 0.84% increase in  $i$ 's likelihood of reporting a YY-chooser oneself, consistent with conditional cooperation.

### *5.5 What accounts for cooperation and defection?*

Although decisions to engage in costly reporting are our focus, we briefly report here to what degree decisions to cooperate or defect are explicable by self-interest given beliefs about others' choices (see Appendix Table B.3 for the details). We also discuss briefly whether the Fehr-Schmidt model can account for instances of cooperation that are inconsistent with self-interest, given beliefs.

A substantial majority of choices between cooperation and defection—63.8 % with the pessimistic 2<sup>nd</sup> game belief, 79.6% with the optimistic belief—are consistent with self-interested rationality. Specifically, assuming each subject's self-reported beliefs to be those on which she in fact conditioned her decision, criterion (1) correctly predicts about 64% of choices between XX and YY, and criterion (1') about 80%, meaning that a selfish rational subject would have chosen as the subject in question did in almost two thirds (80%) of cases, if she were pessimistically (optimistically) assuming that all free second choices by XX-choosers are Y (X). However, while 90% or more of decisions to defect are consistent with self-interest under the pessimistic assumption in each treatment, the proportion of XX choices that are consistent with self-interest is considerably smaller: 26.1% (43.5%), 31.6% (63.2%), 61.9% (90.5%) and 43.5% (78.3%) of cooperators had beliefs making cooperating payoff-maximizing under the pessimistic (optimistic) belief in the HC, MC, LC and NC treatments, respectively. In all, H-SPO is capable

of accounting for many decisions to cooperate, but still leaves either the majority or a substantial minority of those decisions unexplained, depending what the subject was assuming about others' 2<sup>nd</sup> game choices.

Can the latter decisions to cooperate be explained by a simple social preference, e.g. the one used illustratively by us? Leaving details to the Appendix, we find that only a small fraction of the cooperators whose stated beliefs failed to selfishly dictate cooperation under even the optimistic assumption—specifically, about 12%—would have chosen to cooperate out of inequity averse utility maximization given the frequencies of  $\alpha$  and  $\beta$  values assumed by Fehr and Schmidt (2010). Many of the remaining cooperation choices can be explained by aversion to advantageous inequality, but the majority of these cases are consistent with the Fehr-Schmidt model only at degrees of aversion to advantageous relative to disadvantageous inequality that exceed the levels assumed in Fehr and Schmidt's typology. Put differently, the decision-makers in question must strongly wish to avoid defecting against a cooperator and be willing to take a substantial risk of being defected against, equivalent to having  $\beta > \alpha$  in the Fehr-Schmidt model.

## **6. Conclusion**

Numerous studies have focused on the roles social preferences and emotions may play in accounting for decisions favorable to social welfare but lacking benefit to the actor. Surprisingly, the potentially costly act of transmitting information has received little attention. A preference- or emotion-triggered tendency to report at some cost to oneself could be indispensable to social efficiency in situations in which individuals have the option of simply leaving behind undesirable interaction partners without incurring the time, effort or cost to warn others about them. Inclinations to engage in reporting—perhaps closely related to the anthropological and

psychological identification of gossip as a human universal—may be crucial to the viability of many reputation-based incentive systems.

Our experiment underscores the fact that reputation formation may involve costs to those sharing or transmitting the information on which reputational knowledge is based, and it illustrates how laboratory experiments can enhance our understanding of the motivations relevant to decisions to bear such costs. We simplified the problem by assuming that whatever information is transmitted is accurate, and that the recipient knows this to be so with certainty. Although the complications introduced when the message transmitted is a free choice of the reporter and when the recipient must accordingly decide whether to place trust in it require further research, a body of emerging results provides some assurance that most of our findings are likely to hold in such an environment as well.<sup>29</sup>

---

<sup>29</sup> A recent example is Fonseca and Peters (2015), who find that the preponderance of reports about the trustworthiness of trust game players are truthful, and that the availability of such reports raises efficiency. Abraham *et al.* (2015) and Fehr and Sutter (2016) permit false reports but find few, with Abraham *et al.* stating “ratings typically provide a true account” and Fehr and Sutter stating “it turned out that only a small fraction of gossip contains wrong information.” In both experiments, subjects appear to act upon reports presuming their truthfulness.

## REFERENCES

- Ahn, T.K., Justin Esarey and John T. Scholz, 2009, "Reputation and Cooperation in Voluntary Exchanges: Comparing Local and Central Institutions," *Journal of Politics* 71: 398-413.
- Ahn, T.K., Elinor Ostrom, David Schmidt, Robert Shupp and James Walker, 2001, "Cooperation in PD Games: Fear, Greed, and History of Play," *Public Choice* 106: 137-155.
- Bowles, Samuel and Herbert Gintis, 2004, "The Evolution of Strong Reciprocity: Cooperation in Heterogeneous Populations," *Theoretical Population Biology* 65: 17-28.
- \_\_\_\_\_, 2011, *A Cooperative Species: Human Reciprocity and its Evolution*. Princeton, NJ: Princeton University Press.
- Brock, Michelle, Andreas Lange, and Erkut Ozbay, 2013, "Dictating the Risk: Experimental Evidence on Giving in Risky Environments," *American Economic Review* 103(1): 415-437.
- Camerer, Colin, and Richard Thaler, 1995, "Anomalies: Ultimatums, Dictators and Manners," *Journal of Economic Perspectives* 9(2): 209-219.
- Camerer, Colin, 2003, *Behavioral Game Theory: Experiments in Strategic Interaction*. New York: Russell Sage Foundation.
- Camerer, Colin, and George Loewenstein, 2003, "Behavioral Economics: Past, Present, and Future," in C. Camerer, G. Loewenstein and M. Rabin, *Advances in Behavioral Economics*. Princeton, NJ: Princeton University Press.
- Charness, Gary and Matthew Rabin, 2002, "Understanding Social Preferences With Simple Tests," *Quarterly Journal of Economics* 117(3): 817-869.
- Charness, Gary, Luca Rigotti and Aldo Rustichini, 2016, "Social surplus determines cooperation rates in the one-shot Prisoner's Dilemma," *Games and Economic Behavior* 113 – 124.
- Chaudhuri, Ananish, 2011, "Sustaining Cooperation in Laboratory Public Goods Experiments: A Selective Survey of the Literature," *Experimental Economics* 14 (1): 47-83.
- Cooper, Russell, Douglas DeJong, Robert Forsythe, and Thomas Ross, 1996, "Cooperation without Reputation: Experimental Evidence from Prisoner's Dilemma Games," *Games and Economic Behavior* 12: 187-218.

- Cox, James, Daniel Friedman and Steven Gjerstad, 2007, "A Tractable Model of Reciprocity and Fairness," *Games and Economic Behavior* 59 (1): 17-45.
- Dal Bó, Pedro, Andrew Foster and Louis Putterman, 2010, "Institutions and Behavior: Experimental Evidence on the Effects of Democracy," *American Economic Review* 100(5): 2205-2229.
- de Quervain, D.J., Urs Fischbacher, Valerie Treyer, Melanie Schellhammer, Ulrich Schnyder, Alfred Buck, and Ernst Fehr, 2004, "The Neural Basis of Altruistic Punishment," *Science* 305 (5688): 1254-1258.
- Dellarocas, Chrysanthos, 2003, "The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms," *Management Science* 49(10): 1407-1424.
- Dufwenberg, Martin and G. Kirchsteiger, 2004, "A Theory of Reciprocity," *Games and Economic Behavior* 47: 268-298.
- Dunbar, Robin, 2004, "Gossip in Evolutionary Perspective," *Review of General Psychology* 8 (2): 100-110.
- Ertan, Arhan, Talbot Page and Louis Putterman, 2009, "Who to Punish? Individual Decisions and Majority Rule in Mitigating the Free-Rider Problem" *European Economic Review* 53: 495-511.
- Falk, Armin and Urs Fischbacher, 2006, "A Theory of Reciprocity," *Games and Economic Behavior* 54 (2): 293-315.
- Falk, Armin, Ernst Fehr and Urs Fischbacher, 2005, "Driving Forces Behind Informal Sanctions," *Econometrica* 73(6): 2017-2030.
- Fehr, Dietmar and Matthias Sutter, 2016, "Gossip and the Efficiency of Interactions," Working Papers in Economics and Statistics 2016-03.
- Fehr, Ernst and Urs Fischbacher, 2004, "Third-party Punishment and Social Norms," *Evolution and Human Behavior* 25(2): 63-87.
- Fehr, Ernst and Simon Gächter, 2000, "Cooperation and Punishment in Public Goods Experiments," *American Economic Review* 90(4): 980-994.

- Fehr, Ernst and Klaus M. Schmidt, 1999, "A Theory of Fairness, Competition, and Cooperation," *Quarterly Journal of Economics* 114(3): 817-868.
- Fehr, Ernst and Klaus Schmidt, 2010, "On Inequity Aversion: A Reply to Binmore and Shaked," *Journal of Economic Behavior and Organization* 73: 101-108.
- Feinberg, M, Willer, R., Stellar, J., Keltner, D., 2012, "The Virtues of Gossip: Reputational Information Sharing as Prosocial Behavior," *Journal of Personality and Social Psychology*, 102(5): 1015-1030.
- Fischbacher, Urs, 2007. "z-Tree: Zurich Toolbox for Ready-made Economic Experiments," *Experimental Economics* 10: 171-178.
- Fischbacher, Urs and Simon Gächter, 2010, "Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Good Experiments," *American Economic Review* 100(1): 541-56.
- Fonseca, Miguel and Kim Peters, "Gossip among Trustors Increases Trust and Trustworthiness in the Trust Game," paper presented at Economic Science Association World Meetings, Sydney, Australia, July 2015.
- Gächter, Simon and Benedikt Herrmann, 2009, "Reciprocity, Culture and Human Cooperation: Previous Insights and a New Cross-Cultural Experiment," *Philosophical Transactions of the Royal Society B* 364(1518): 791-806.
- Gërxxhani, Klarita, Jordi Brandts and Arthur Schram, 2013, "The Emergence of Employer Information Networks in an Experimental Labor Market," *Social Networks* 35: 541-560.
- Gintis, Herbert, Samuel Bowles, Robert Boyd and Ernst Fehr, eds., 2005, *Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life*. Cambridge: MIT Press.
- Gregg, Dawn G., and Judy E. Scott, 2006, "The role of reputation systems in reducing online auction fraud," *International Journal of Electronic Commerce* 10.3: 95-120.
- Healy, Paul, 2007, "Group Reputations, Stereotypes, and Cooperation in a Repeated Labor Market," *American Economic Review* 97(5): 1751-1773.
- Hoffman, Elizabeth, Kevin McCabe and Vernon Smith, 1998, "Behavioral Foundations of Reciprocity: Experimental Economics and Evolutionary Psychology," *Economic Inquiry* 36: 335-352.

- Jackson, Matthew and Yves Zenou, 2014, "Games on Networks," in Peyton Young and Shmuel Zamir, eds., *Handbook of Game Theory, Vol. 4*. Amsterdam: Elsevier.
- Kamei, Kenju, 2014, "Conditional Punishment," *Economics Letters* 124: 199-202.
- Kamei, Kenju, forthcoming, "Promoting Competition or Helping the Less Endowed? Distributional Preferences and Collective Institutional Choices under Intra-Group Inequality," *Journal of Conflict Resolution* (in press).
- Kamei, Kenju and Louis Putterman, forthcoming, "Play it Again: Partner Choice, Reputation Building and Learning from Finitely-Repeated Dilemma Games," *Economic Journal* (in press).
- Kreps, David, Paul Milgrom, John Roberts and Robert Wilson, 1982, "Rational Cooperation in Finitely Repeated Prisoners' Dilemma," *Journal of Economic Theory* 27: 245-252.
- Loewenstein, George, 2000, "Emotions in Economic Theory and Economic Behavior," *American Economic Review* 90(2): 426-432.
- Palfrey, Thomas and Jeffrey Prisbrey, 1997, "Anomalous Behavior in Public Goods Experiments: How Much and Why?" *American Economic Review* 87(5): 829-846.
- Pinker, Steven, 2003, *The Blank Slate: The Modern Denial of Human Nature*. New York: Viking.
- Resnick, Paul and Richard Zeckhauser, 2002, "Trust among strangers in Internet transactions: Empirical analysis of eBay's reputation system," *Advances in applied microeconomics* 11: 127-157.
- Saito, Kota, 2013, "Social Preferences under Risk: Equality of Opportunity versus Equality of Outcome," *American Economic Review* 103(7): 3084-3101.
- Shampanier, Kristina, Nina Mazar, Dan Ariely, 2007, "Zero as a Special Price: The True Value of Free Products," *Marketing Science* 26: 742-757.
- Smith, Adam, 1761. *The Theory of Moral Sentiments* (2<sup>nd</sup> ed.). Strand and Edinburgh: A. Millar.
- Sobel, Joel, 2005, "Interdependent Preferences and Reciprocity," *Journal of Economic Literature*, XLIII: 392-436.



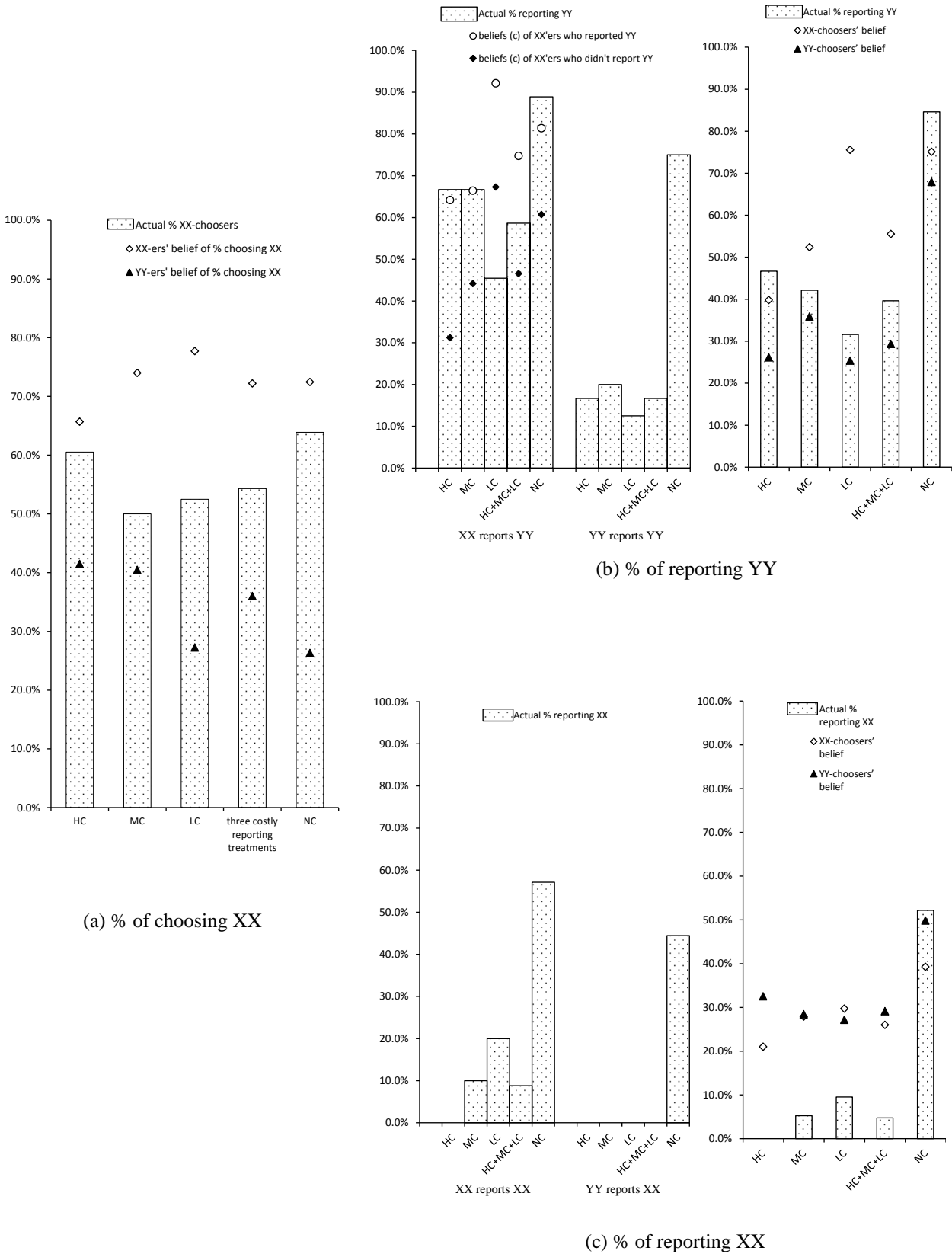
Wang, Zhongmin, 2010, "Anonymity, social image, and the competition for volunteers: A case study of the online market for reviews," *The BE Journal of Economic Analysis & Policy* 10(1): 1935-1682.

Wilson, Edward O., 2012, *The Social Conquest of Earth*. New York: Liveright.

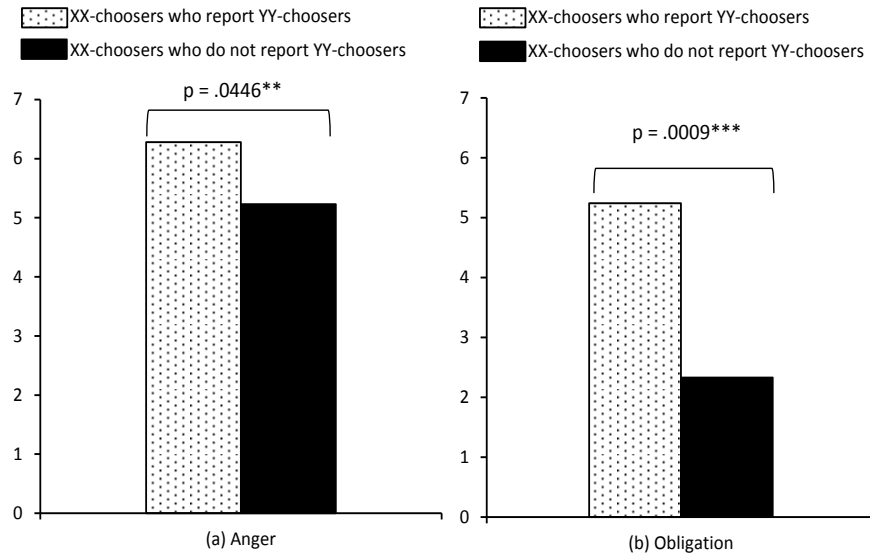
**FIGURE 1:** Payoff Matrix

		Player 2	
		X	Y
Player 1	X	\$10, \$10	\$4, \$11
	Y	\$11, \$4	\$5, \$5

FIGURE 2: Choices of XX or YY, Reporting Decisions, and Beliefs



**FIGURE 3: Average Anger Level and Average Feeling of Obligation to Help Third Party  $k$**



*Notes:*  $p$ -values indicate Mann-Whitney-test results (two-sided). The Anger variable is a subject's response to the following question: "How did you feel about your first counterpart's decision? Please rate on a scale from 1 = very pleased to 7 = very angry." The Obligation variable is a subject's response to the following question: "Did you feel a sense of obligation to help your first counterpart's next counterpart by sending a report? Please rate on a scale from 1 = did not feel obligated at all, to 7 = felt strongly obligated."

\*\*\* Significant at the 1 percent level.

\*\* Significant at the 5 percent level.

\* Significant at the 10 percent level.

## Not for Publication

Supplementary Online Appendix for Kamei and Putterman,  
“Reputation Transmission without Benefit to the Reporter:  
a Behavioral Underpinning of Markets in Experimental Focus”

Kenju Kamei and Louis Putterman<sup>1</sup>

### Table of Contents

Appendix A: Theoretical Analysis .....	p. 2
Appendix B: Additional Tables and Figures .....	p. 23
Appendix C: Instructions for the LC treatment .....	p. 38

---

<sup>1</sup> Department of Economics and Finance, Durham University (email: kenju.kamei@durham.ac.uk, kenju.kamei@gmail.com) and Department of Economics, Brown University (email: Louis\_Putterman@brown.edu), respectively.

## Appendix A: Theoretical Analysis

### A.1. Deriving Condition (1)

Under the “pessimistic” assumption, decision-maker  $i$  assumes that an individual free to revise her second choice always selects Y. As shown in TABLE A.1, the combined expected payoffs of subject  $i$  from cooperating or defecting, respectively, in the two interactions, can then be expressed as

$$E[\pi_i(XX)] = 8 + 12a_i - 5a_ib_i + c_i - a_ic_i (= 4 + 6a_i + 4 + 6a_i - 5a_ib_i + c_i - a_ic_i)$$

and

$$E[\pi_i(YY)] = 10 + 12a_i - 6a_ic_i (= 5 + 6a_i + 5 + 6a_i - 6a_ic_i).$$

Since a selfish player would choose XX (YY) if  $E[\pi_i(XX)] > (<) E[\pi_i(YY)]$ , we obtain the payoff-maximization condition:

$$\text{cooperate (defect) if } 5a_i(c_i - b_i) + c_i > (<) 2. \quad (1)$$

Player  $i$  is indifferent between cooperating and defecting if  $5a_i(c_i - b_i) + c_i = 2$ .

**TABLE A.1:** Expected Payoff for Each of Initial Choices (XX or YY)

## (1) Expected Payoffs in the First Interaction

Own Decision	First Partner's Decision	Probability	Payoff	Expected Payoff
XX	XX	$a_i$	10	$4 + 6a_i$
	YY	$1 - a_i$	4	
YY	XX	$a_i$	11	$5 + 6a_i$
	YY	$1 - a_i$	5	

## (2) Expected Payoffs in the Second Interaction

Initial Decision	Second Partner's Decision	Player Receiving Report	Outcome	Probability	Payoff	Expected Payoff
XX	X	None	X, X	$a_i(1-b_i)^2$	10	$4 + 6a_i - 5a_ib_i + c_i - a_ic_i$
		Only subject	Y, X	$a_ib_i(1-b_i)$	11	
		Only partner	X, Y	$a_ib_i(1-b_i)$	4	
		Both	Y, Y	$a_ib_i^2$	5	
	Y	None	X, Y	$(1-a_i)(1-b_i)(1-c_i)$	4	
		Only subject	Y, Y	$(1-a_i)(1-b_i)c_i$	5	
		Only partner	X, Y	$(1-a_i)b_i(1-c_i)$	4	
		Both	Y, Y	$(1-a_i)b_ic_i$	5	
YY	X	None	Y, X	$a_i(1-b_i)(1-c_i)$	11	$5 + 6a_i - 6a_ic_i$
		Only subject	Y, X	$a_ib_i(1-c_i)$	11	
		Only partner	Y, Y	$a_i(1-b_i)c_i$	5	
		Both	Y, Y	$a_ib_ic_i$	5	
	Y	None				
		Only subject	Y, Y	$1 - a_i$	5	
		Only partner				
		Both				

Notes:  $a_i$  indicates the fraction of subjects in the session that subject  $i$  believes will select XX.  $b_i$  indicates the fraction of subjects she believes will engage in reporting those who select XX.  $c_i$  indicates the fraction of subjects she believes will engage in reporting those who select YY.

**A.2. Revising Condition (1) Based on the Optimistic Belief that Cooperators Cooperate with Cooperators given Free 2<sup>nd</sup> Choices**

Subject  $i$  chooses XX if  $E[\pi_i(XX)] > E[\pi_i(YY)]$ , i.e.,

$$4 + 6a_i + 4 + 6a_i + a_i b_i + c_i - a_i c_i > 5 + 6a_i + 5 + 6a_i - 6a_i c_i.$$

That is,  $5a_i c_i + a_i b_i + c_i > 2$ .

This means the higher is subject  $i$ 's belief  $a_i$ , the more likely subject  $i$  is to choose XX. Also, we learn that the higher is subject  $i$ 's belief  $c_i$ , the more likely subject  $i$  is to choose XX. The expected payoffs of subject  $i$  in the first interaction are as in Panel (1) of TABLE 1. Those in the second interaction are summarized in TABLE A.1 below.



**TABLE A.2:** Revised Expected Material Payoffs in the Second Interaction using optimistic assumption about XX-choosers' free 2<sup>nd</sup> choices

Own Initial Decision	Second Partner's Decision	Player Receiving Report	Outcome	Probability	Payoff	Expected Payoff	
XX	X	None	X, X	$a_i(1-b_i)^2$	10	$4 + 6a_i + a_i b_i + c_i - a_i c_i$	
		Only subject	Y, X	$a_i b_i(1-b_i)$	11		
		Only partner	X, X	$a_i b_i(1-b_i)$	10		
		Both	Y, X	$a_i b_i^2$	11		
	Y	None	X, Y	$(1-a_i)(1-b_i)(1-c_i)$	4		
		Only subject	Y, Y	$(1-a_i)(1-b_i)c_i$	5		
		Only partner	X, Y	$(1-a_i)b_i(1-c_i)$	4		
		Both	Y, Y	$(1-a_i)b_i c_i$	5		
YY	X	None	Y, X	$a_i(1-b_i)(1-c_i)$	11	$5 + 6a_i - 6a_i c_i$	
		Only subject	Y, X	$a_i b_i(1-c_i)$	11		
		Only partner	Y, Y	$a_i(1-b_i)c_i$	5		
		Both	Y, Y	$a_i b_i c_i$	5		
	Y	None					
		Only subject	Y, Y	$1 - a_i$	5		
		Only partner					
		Both					

Notes:  $a_i$  indicates the fraction of subjects in the session that subject  $i$  believes will select XX.  $b_i$  indicates the fraction of subjects she believes will engage in reporting those who select XX.  $c_i$  indicates the fraction of subjects she believes will engage in reporting those who select YY. The decision-maker  $i$  performing this calculation is self-interested and hence chooses Y when given a free 2<sup>nd</sup> choice, but  $i$  “optimistically” assumes that others who select XX initially will choose X in a free 2<sup>nd</sup> choice if meeting another XX chooser; that is, all XX-choosers other than (possibly)  $i$  are assumed to have a social preference or other reason for choosing X in their 2<sup>nd</sup> interaction, if they meet another cooperator.

### A.3. Conditions under which an Inequity-Averse Subject $i$ Chooses XX.

If we assume that a subject  $i$  is an inequity-averse agent, her payoff matrix is expressed as below:

**Figure:** Amended Payoff Matrix incorporating Inequity-Averse Preferences into FIGURE 1 of the paper.

		Subject $j$	
		$X \text{ with prob. } a_i$	$Y \text{ with prob. } 1 - a_i$
Subject $i$	X	$\underline{10 + E[XX]}, 10 + E[XX]$	$\underline{4 + E[XX] - \alpha_i \cdot (11 + E[YY] - 4 - E[XX])},$ $11 + E[YY] - \beta_j \cdot (11 + E[YY] - 4 - E[XX])$
	Y	$\underline{11 + E[YY] - \beta_j \cdot (11 + E[YY] - 4 - E[XX])},$ $4 + E[XX] - \alpha_i \cdot (11 + E[YY] - 4 - E[XX])$	$\underline{5 + E[YY]}, 5 + E[YY]$

*Note:* The underlined payoffs are the payoffs of subject  $i$ . We assume that subject  $i$  considers inequality with her first-interaction partner (subject  $j$ ) only. Subject  $i$  believes that she meets with a cooperator with a probability  $a_i$ .

In the payoff matrix,  $E[XX]$  ( $E[YY]$ ) is the expected payoff in the second period when choosing XX (YY) based on subject  $i$ 's beliefs. Under the assumption of pessimistic beliefs,  $E[XX]$  and  $E[YY]$  are given by Panel (2) of TABLE A.1. Under the assumption of optimistic beliefs,  $E[XX]$  and  $E[YY]$  are given by TABLE A.2.

The expected payoff of subject  $i$  when choosing XX is:

$$\pi(XX) = a_i \cdot (10 + E[XX]) + (1 - a_i) \cdot (4 + E[XX] - \alpha_i \cdot (11 + E[YY] - 4 - E[XX])).$$

By contrast, the expected payoff of subject  $i$  when choosing YY is:

$$\pi(YY) = a_i \cdot (11 + E[YY] - \beta_j \cdot (11 + E[YY] - 4 - E[XX])) + (1 - a_i) \cdot (5 + E[YY]).$$

Subject  $i$  chooses XX (YY) if  $\pi(XX) > (<) \pi(YY)$ . In other words,

$$\begin{aligned} \pi(XX) - \pi(YY) \\ = E[XX] - E[YY] - 1 + (7 + E[YY] - E[XX]) \cdot (a_i \beta_j - \alpha_i (1 - a_i)) > (<) 0. \end{aligned} \quad (A1)$$

Case 1: The Pessimistic Beliefs

From condition (A1) and Panel (2) of TABLE 1, we have: subject  $i$  chooses  $XX$  ( $YY$ ) if and only if

$$5a_i \cdot (c_i - b_i) + c_i > (<) 2 - (8 - 5a_i \cdot (c_i - b_i) - c_i) \cdot (a_i \cdot \beta_i - (1 - a_i) \cdot \alpha_i) \quad (A2)$$

Case 2: The Optimistic Beliefs

From condition (A1) and TABLE A.1, we have: subject  $i$  chooses  $XX$  ( $YY$ ) if and only if

$$5a_i \cdot c_i + a_i b_i + c_i > (<) 2 - (8 - 5a_i \cdot c_i - a_i b_i - c_i) \cdot (a_i \beta_i - (1 - a_i) \alpha_i) \quad (A3)$$

#### A.4. Reporting Decisions of an Inequity-Averse Subject

##### A.4.1. Case 1 – A cooperator meets another cooperator in period 1

PROPOSITION A1: *Suppose that  $\beta < 1$ . Also suppose that a cooperator  $i$  forms the pessimistic belief. Then, the cooperator reports (does not report) her matched cooperator if and only if the following two conditions hold (do not hold):*

$$6a_i > \rho \text{ and } -\rho + ((\alpha_i + \beta_i)b_i - \beta_i)(6a_i - \rho) > 0.$$

*Suppose instead that the cooperator forms the optimistic belief. Then, the cooperator never reports her matched cooperator.*

Proof:

Suppose that subject  $i$  has been matched with subject  $j$  in period 1 and both subjects select XX. Then, subjects  $i$  and  $j$  each receive a payoff of 10 points in that period. We examine the conditions under which cooperator  $i$  reports her matched cooperator  $j$ . We consider the two assumptions on  $i$ 's beliefs.

(a) Suppose that cooperator  $i$  forms the pessimistic belief. Under this assumption, from TABLE A.1,  $i$ 's total expected payoff in the experiment is calculated as:

$$\pi_i(XX) = 14 - \rho \cdot 1_{report} + 6a_i - 5a_i b_i + c_i - a_i c_i.^2 \quad (A7)$$

By contrast,  $i$ 's belief about  $j$ 's material payoff is dependent on whether  $i$  reports  $j$  or not:

(i) If  $i$  reports  $j$ :

$$\pi_j|reported = 10 - \rho b_i + (\text{Expected payoff of } j \text{ in period 2 if } i \text{ reports } j)$$

*Expected payoff of  $j$  in period 2 if  $i$  reports  $j$*

$$\begin{aligned} &= a_i b_i (5) + a_i (1 - b_i) (4) + (1 - a_i) (c_i) (5) + (1 - a_i) (1 - c_i) (4) \\ &= 4 + a_i b_i + c_i - a_i c_i. \end{aligned}$$

This expected payoff is calculated based on TABLE A.3 below.

Thus, we have:

$$\pi_j|reported = 14 - \rho b_i + a_i b_i + c_i - a_i c_i. \quad (A8)$$

---

<sup>2</sup>  $1_{report} = 1$  if  $i$  reports  $j$ ; 0 otherwise.  $\rho = 1, 0.5$  and  $0.05$  for the HC, MC and LC treatments, respectively.

TABLE A.3: Cooperator  $i$ 's Belief about Cooperator  $j$ 's Expected Material Payoff in Period 2 if  $i$  Reports  $j$  (Pessimistic Assumption)

$j$ 's Initial Decision	Second Partner's Decision	Recipient of Report	Outcome	Probability	Payoff	Expected Payoff
XX	X	Only $j$ 's next partner	X,Y	$a_i(1-b_i)$	4	$4+a_i b_i + c_i - a_i c_i$
		Both $j$ and $j$ 's next partner	Y,Y	$a_i b_i$	5	
	Y	Only $j$ 's next partner	X,Y	$(1-a_i)(1-c_i)$	4	
		Both $j$ and $j$ 's next partner	Y,Y	$(1-a_i)c_i$	5	

(ii) If  $i$  does not report  $j$ :

$$\pi_j|not\ reported = 10 - \rho b_i + (\text{Expected payoff of } j \text{ in period 2 if } i \text{ does not report } j)$$

$$\text{Expected payoff of } j \text{ in period 2 if } i \text{ does not report } j = 4 + 6a_i + a_i b_i + c_i - a_i c_i.$$

This expected payoff in period 2 is calculated based on TABLE A.4 below.

Thus, we have:

$$\pi_j|not\ reported = 14 - \rho b_i + 6a_i + a_i b_i + c_i - a_i c_i. \quad (A9)$$

TABLE A.4: Cooperator  $i$ 's Belief about Cooperator  $j$ 's Expected Material Payoff in Period 2 if  $i$  Does Not Report  $j$  (Pessimistic Assumption)

$j$ 's Initial Decision	Second Partner's Decision	Recipient of Report	Outcome	Probability	Payoff	Expected Payoff
XX	X	None	X,X	$a_i(1-b_i)$	10	$4+6a_i+a_i b_i - a_i c_i + c_i$
		Only $j$	Y,X	$a_i b_i$	11	
	Y	None	X,Y	$(1-a_i)(1-c_i)$	4	
		Only $j$	Y, Y	$(1-a_i)c_i$	5	

From equations (A7), (A8) and (A9), we find that if cooperator  $i$  reports cooperator  $j$ ,  $i$ 's material payoff ( $j$ 's material payoff) is bigger if  $6a_i > \rho$  ( $6a_i < \rho$ ). To see this:

$$\begin{aligned} \pi_i|report - \pi_j|reported &= 6(a_i - a_i b_i) - \rho(1 - b_i) \\ &= (6a_i - \rho)(1 - b_i), \end{aligned}$$

which is positive (negative) if  $6a_i > \rho$  ( $6a_i < \rho$ ).

If cooperator  $i$  does not report cooperator  $j$ ,  $i$ 's material payoff ( $j$ 's material payoff) is bigger if  $6a_i < \rho$  ( $6a_i > \rho$ ). To see this,

$$\pi_i|not\ report - \pi_j|not\ reported = \rho b_i - 6a_i b_i = (\rho - 6a_i)b_i,$$

which is positive (negative) if  $6a_i < \rho$  ( $6a_i > \rho$ ).

In summary, we need to consider the following two possible situations:

- (i)  $6a_i < \rho$  ( $j$ 's material payoff is bigger if  $i$  reports  $j$ ;  $i$ 's material payoff is bigger if  $i$  does not report  $j$ )
- (ii)  $6a_i > \rho$  ( $i$ 's material payoff is bigger if  $i$  reports  $j$ ;  $j$ 's material payoff is bigger if  $i$  does not report  $j$ )

Suppose first that  $6a_i < \rho$ . Then, cooperator  $i$ 's Fehr-Schmidt expected utility when  $i$  chooses to report is:

$$\begin{aligned} u(\pi_i|\pi_j) |report &= \pi_i |report - \alpha_i \max\{\pi_j |reported - \pi_i |report, 0\} \\ &= 14 - \rho + 6a_i - 5a_i b_i + c_i - a_i c_i - \alpha_i (\rho - \rho b_i - 6a_i + 6a_i b_i). \end{aligned}$$

By contrast,  $i$ 's Fehr-Schmidt expected utility when  $i$  chooses not to report is:

$$\begin{aligned} u(\pi_i|\pi_j) |not report &= \pi_i |not report - \beta_i \max\{\pi_i |not report - \pi_j |not reported, 0\} \\ &= 14 + 6a_i - 5a_i b_i + c_i - a_i c_i - \beta_i (\rho b_i - 6a_i b_i). \end{aligned}$$

In other words,  $i$  will report  $j$  if and only if:

$$\begin{aligned} -\rho - \alpha_i (\rho - \rho b_i - 6a_i + 6a_i b_i) &> -\beta_i (\rho b_i - 6a_i b_i), \text{ or} \\ -\rho + [\beta_i b_i - \alpha_i (1 - b_i)](\rho - 6a_i) &> 0. \end{aligned}$$

This condition does not hold as we are assuming that  $0 \leq \beta_i \leq \alpha_i < 1$ . This is because:

$$\begin{aligned} \text{the left-hand side} &= -\rho + [\beta_i b_i - \alpha_i (1 - b_i)](\rho - 6a_i) < -\rho + [\beta_i b_i - \alpha_i (1 - b_i)](\rho - 6a_i)|_{\alpha_i=0, \beta_i=1} \\ &= -\rho + b_i(\rho - 6a_i) \\ &< -\rho + b_i(\rho - 6a_i)|_{b_i=1} = -6a_i. \end{aligned}$$

Suppose instead that  $6a_i > \rho$ . Then, cooperator  $i$ 's Fehr-Schmidt expected utility when  $i$  chooses to report is calculated as:

$$\begin{aligned} u(\pi_i|\pi_j) |report &= \pi_i |report - \beta_i \max\{\pi_i |report - \pi_j |reported, 0\} \\ &= 14 - \rho + 6a_i - 5a_i b_i + c_i - a_i c_i - \beta_i (-\rho + \rho b_i + 6a_i - 6a_i b_i). \end{aligned}$$

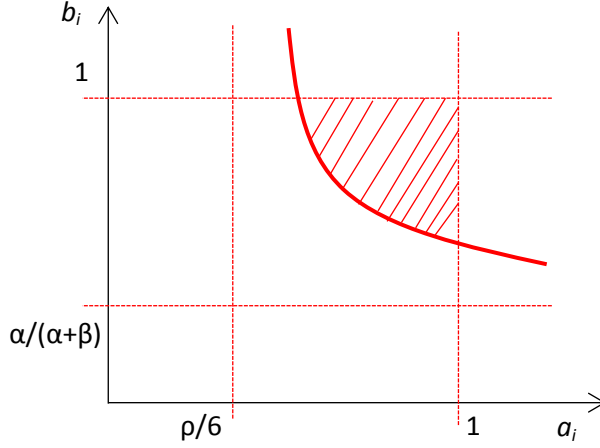
By contrast,  $i$ 's Fehr-Schmidt expected utility when  $i$  chooses not to report is expressed as:

$$\begin{aligned} u(\pi_i|\pi_j) |not report &= \pi_i |not report - \alpha_i \max\{\pi_j |not reported - \pi_i |not report, 0\} \\ &= 14 + 6a_i - 5a_i b_i + c_i - a_i c_i - \alpha_i (6a_i b_i - \rho b_i). \end{aligned}$$

In other words,  $i$  will report  $j$  if and only if:

$$\begin{aligned} -\rho - \beta_i (-\rho + \rho b_i + 6a_i - 6a_i b_i) &> -\alpha_i (6a_i b_i - \rho b_i), \text{ or} \\ -\rho + (-\beta_i + \beta_i b_i + \alpha_i b_i)(6a_i - \rho) &> 0. \end{aligned} \tag{A10}$$

This suggests that the higher  $b$  cooperator  $i$  has, the more likely  $i$  is to report  $j$ .



	LHS of Condition (A10)	Condition of $b$ for a cooperator to report her matched cooperator, given $a$ and $\rho$ .
$\alpha_i = 0, \beta_i = 0$	$-\rho$	Reporting never happens
$\alpha_i = 0.5, \beta_i = 0.25$	$-\rho + (-.25 + .75b_i)(6a_i - \rho)$	$b_i > 1/3 + (4/3) \cdot \rho / (6a_i - \rho)$
$\alpha_i = 1, \beta_i = 0.6$	$-\rho + (-.6 + 1.6b_i)(6a_i - \rho)$	$b_i > 3/8 + (5/8) \cdot \rho / (6a_i - \rho)$

(b) Suppose next that cooperator  $i$  forms the optimistic belief. Under this assumption, from TABLE A.2,  $i$ 's total expected payoff is calculated as:

$$\pi_i(XX) = 10 - \rho \cdot 1_{report} + (\text{Expected payoff of } i \text{ in period 2}),$$

$$\text{where Expected payoff of } i \text{ in period 2} = 4 + 6a_i + a_i b_i + c_i - a_i c_i.$$

Thus, we have:

$$\pi_i(XX) = 14 - \rho \cdot 1_{report} + 6a_i + a_i b_i + c_i - a_i c_i.$$

Cooperator  $i$ 's belief about  $j$ 's material payoff depends on whether  $i$  reports  $j$  or not:

(a) If  $i$  reports  $j$ :

$$\pi_j|reported = 10 - \rho b_i + (\text{Expected payoff of } j \text{ in period 2 if } i \text{ reports } j),$$

$$\text{where Expected payoff of } j \text{ in period 2 if } i \text{ reports } j = 4 + 6a_i + c_i - a_i c_i.$$

Here, the expected payoff of  $j$  in period 2 is calculated based on TABLE A.5.

Thus we have:

$$\pi_j|reported = 14 - \rho b_i + 6a_i + c_i - a_i c_i.$$

TABLE A.5: Cooperator  $i$ 's Belief about Cooperator  $j$ 's Expected Material Payoff in Period 2 if  $i$  Reports  $j$  (Optimistic Assumption)

$j$ 's Initial Decision	Second Partner's Decision	Recipient of Report	Outcome	Probability	Payoff	Expected Payoff
XX	X	Only $j$ 's next partner	X,X	$a_i(1-b_i)$	10	$4+6a_i+c_i - a_i c_i$
		Both $j$ and $j$ 's next partner	X,X	$a_i b_i$	10	
	Y	Only $j$ 's next partner	X,Y	$(1-a_i)(1-c_i)$	4	
		Both $j$ and $j$ 's next partner	Y, Y	$(1-a_i)c_i$	5	

(b) If  $i$  does not report  $j$ :

$$\pi_j | \text{not reported} = 10 - \rho b + (\text{Expected payoff of } j \text{ in period 2 if } i \text{ does not report } j),$$

$$\text{where Expected payoff of } j \text{ in period 2 if } i \text{ does not report } j = 4 + 6a_i + c_i - a_i c_i.$$

Thus we have:

$$\pi_i | \text{not reported} = 14 - \rho b_i + 6a_i + c_i - a_i c_i.$$

TABLE A.6: Cooperator  $i$ 's Belief about Cooperator  $j$ 's Expected Material Payoff in Period 2 if  $i$  Does Not Report  $j$  (Optimistic Assumption)

$j$ 's Initial Decision	Second Partner's Decision	Recipient of Report	Outcome	Probability	Payoff	Expected Payoff
XX	X	None	X,X	$a_i(1-b_i)$	10	$4+6a_i+c_i - a_i c_i$
		Only $j$	X,X	$a_i b_i$	10	
	Y	None	X,Y	$(1-a_i)(1-c_i)$	4	
		Only $j$	Y, Y	$(1-a_i)c_i$	5	

Regarding income inequality between  $i$  and  $j$ , if  $i$  reports  $j$ ,  $i$ 's material payoff is larger than  $j$ 's if:

$$\begin{aligned} \pi_i | \text{report} &> \pi_j | \text{reported}, \text{ or} \\ -\rho + a_i b_i &> -\rho b_i, \text{ or} \\ a_i b_i &> \rho(1 - b_i). \end{aligned} \tag{A11}$$

By contrast, if  $i$  does not report  $j$ ,  $i$ 's material payoff is bigger than  $j$ 's if:

$$\begin{aligned} \pi_i | \text{not report} &> \pi_j | \text{not reported}, \text{ or} \\ a_i b_i &> -\rho b_i, \text{ or} \\ a_i b_i + \rho b_i &> 0. \end{aligned} \tag{A12}$$



Condition (A12) always holds unless  $i$ 's beliefs (and/or  $\rho$ ) are altogether zero. Thus, we only need to consider two cases:  $a_i b_i > \rho(1 - b_i)$  and  $a_i b_i < \rho(1 - b_i)$ .

Suppose first that  $a_i b_i > \rho(1 - b_i)$ . Whether  $i$  reports  $j$  or not,  $i$ 's material payoff is bigger than  $j$ 's. In this situation, cooperator  $i$ 's Fehr-Schmidt expected utility when  $i$  chooses to report  $j$  is:

$$\begin{aligned} u(\pi_i | \pi_j) | \text{report} &= \pi_i | \text{report} - \beta_i \max\{\pi_i | \text{report} - \pi_j | \text{reported}, 0\} \\ &= 14 - \rho + 6a_i + a_i b_i + c_i - a_i c_i - \beta_i(-\rho + a_i b_i + \rho b_i). \end{aligned}$$

By contrast, cooperator  $i$ 's Fehr-Schmidt expected utility when  $i$  chooses not to report is:

$$\begin{aligned} u(\pi_i | \pi_j) | \text{not report} &= \pi_i | \text{report} - \beta_i \max\{\pi_i | \text{not report} - \pi_j | \text{not reported}, 0\} \\ &= 14 + 6a_i + a_i b_i + c_i - a_i c_i - \beta_i(a_i b_i + \rho b_i). \end{aligned}$$

We see that  $i$  does not report  $j$  as we are assuming that  $\beta_i < 1$ . This is because:

$$u(\pi_i | \pi_j) | \text{report} - u(\pi_i | \pi_j) | \text{not report} = -\rho - \beta_i(-\rho) = \rho(\beta_i - 1) < 0.$$

Suppose instead that  $a_i b_i < \rho(1 - b_i)$ . In this situation, cooperator  $i$ 's Fehr-Schmidt expected utility when  $i$  chooses to report  $j$  is:

$$\begin{aligned} u(\pi_i | \pi_j) | \text{report} &= \pi_i | \text{report} - \alpha_i \max\{\pi_j | \text{reported} - \pi_i | \text{report}, 0\} \\ &= 14 - \rho + 6a_i + a_i b_i + c_i - a_i c_i - \alpha_i(-\rho b_i + \rho - a_i b_i). \end{aligned}$$

Thus, in this case,  $i$  reports  $j$  if and only if:

$$\begin{aligned} u(\pi_i | \pi_j) | \text{report} - u(\pi_i | \pi_j) | \text{not report} &> 0, \text{ or} \\ -\rho - \alpha_i(-\rho b_i + \rho - a_i b_i) + \beta_i(a_i b_i + \rho b_i) &> 0, \text{ or} \\ (\alpha_i + \beta_i)a_i b_i - (1 + \alpha_i)\rho + \rho b_i(\alpha_i + \beta_i) &> 0, \text{ or} \\ (\alpha_i + \beta_i)(a_i b_i + \rho b_i) &> (1 + \alpha_i)\rho. \end{aligned}$$

Suppose that  $0 \leq \beta_i < 1$ . Then, this condition implies:

$$(\alpha_i + \beta_i)(a_i b_i + \rho b_i) > (1 + \alpha_i)\rho > (\beta_i + \alpha_i)\rho,$$

which means that:

$$a_i b_i > \rho(1 - b_i).$$

This cannot be held as we are assuming that  $a_i b_i < \rho(1 - b_i)$ .

In other words, cooperator  $i$  never reports  $j$  if  $0 \leq \beta_i < 1$  and  $i$  forms the optimistic belief.  $\square$

**COROLLARY A1:** *Suppose that  $\beta_i < 1$ . Then, subject  $i$  engages in more costly reporting in Case 2 ( $i$  chooses  $XX$ ; and then meets with a defector) than in Case 1.*

Proof:

As discussed in the manuscript, when a cooperator  $i$  meets with a defector  $j$ ,  $i$  reports  $j$  if and only if:

$$a_i > \frac{\rho}{6} + \frac{\rho}{6\alpha_i}.$$

This condition is obtained by re-arranging Equation (3) of the manuscript. Recall that this condition holds regardless of which assumption, pessimistic or optimistic, we impose.

From Proposition A1, if cooperator  $i$  forms the pessimistic belief and meets with another cooperator  $k$ ,  $i$  reports  $k$  if and only if:

$$6a_i > \rho \text{ and } -\rho + ((\alpha_i + \beta_i)b_i - \beta_i)(6a_i - \rho) > 0.$$

Here, we see that an additional requirement that  $i$  will report  $k$  is:  $b_i > \frac{\beta_i}{\alpha_i + \beta_i}$ ; otherwise,  $-\rho + ((\alpha_i + \beta_i)b_i - \beta_i)(6a_i - \rho) \leq 0$ .

The condition of  $-\rho + ((\alpha_i + \beta_i)b_i - \beta_i)(6a_i - \rho) > 0$  in Proposition A1 is stronger than  $a_i > \frac{\rho}{6} + \frac{\rho}{6\alpha_i}$ . To see this, re-arranging the condition:  $-\rho + ((\alpha_i + \beta_i)b_i - \beta_i)(6a_i - \rho) > 0$ , we obtain:

$$a_i > \frac{\rho}{6} + \frac{\rho}{6[(\alpha_i + \beta_i)b_i - \beta_i]},$$

whose right hand side is greater than or equal to:  $\frac{\rho}{6} + \frac{\rho}{6[(\alpha_i + \beta_i)b_i - \beta_i]}|_{b_i=1} = \frac{\rho}{6} + \frac{\rho}{6\alpha_i}$ . This means that costly reporting is more likely to be realized in Case 2 than in Case 1 under the pessimistic assumption.

Suppose instead that cooperator  $i$  forms the optimistic belief. Then, when  $i$  meets with a cooperator  $k$ ,  $i$  will never report  $k$  (as shown in Proposition A1), unlike the situation in which  $i$  meets with a defector.

□

#### A.4.2. Case 2 – A cooperator meets a defector in period 1.

PROPOSITION A2: *Regardless of which belief, either pessimistic or optimistic, a cooperator  $i$  forms, a cooperator reports (does not report) her matched defector if and only if the following condition hold (do not hold):*

$$-\rho + (6a_i - \rho)a_i > 0$$

Proof:

Suppose that subject  $i$  chose XX and has been matched with subject  $j$  who selected YY in period 1. Then, cooperator  $i$  received a payoff of 4 points in that period; defector  $j$  received a payoff of 11 points in that period. We examine the conditions under which cooperator  $i$  reports his matched defector  $j$ .

Suppose that subject  $i$  has the utility function defined in Eq. (2). Subject  $i$  reports subject  $j$  if and only if  $u_i(\pi_i | \pi_j) | \text{report} > u_i(\pi_i | \pi_j) | \text{not report}$  (i.e.,  $i$ 's utility when  $i$  reports  $j$  is greater than  $i$ 's utility when  $i$  does not report  $j$ ).

The cooperator  $i$ 's totally expected payoff is:

$$\begin{aligned} \pi_i &= (\text{Period 1 payoff}) - \rho \cdot 1_{\text{report}} + (\text{Expected payoff in period 2}) \\ &= 4 + (\text{Expected payoff in period 2}) - \rho \cdot 1_{\text{report}}. \end{aligned} \quad (\text{A13})$$

By contrast, cooperator  $i$ 's belief about defector  $j$ 's material payoff is dependent on whether  $i$  reports  $j$  or not:

(a) When  $i$  reports  $j$ :

$$\begin{aligned} \pi_j | \text{reported} &= (\text{Period 1 payoff}) - \rho b_i + (\text{Expected payoff in period 2 if } i \text{ reports } j) \\ &= 11 - \rho b_i + 5 \\ &= 16 - \rho b_i. \end{aligned} \quad (\text{A14})$$

(b) When  $i$  does not report  $j$ :

$$\begin{aligned} \pi_j | \text{not reported} &= (\text{Period 1 payoff}) - \rho \cdot b_i + (\text{Expected Payoff in period 2 if } i \text{ does not report } j) \\ &= 11 - \rho b_i + 11 \cdot a_i + 5 \cdot (1 - a_i) \\ &= 16 + 6a_i - \rho b_i. \end{aligned} \quad (\text{A15})$$

Here, the utility of subject  $i$  is calculated using (A13) and (A14) or (A15).

(i) The utility of cooperator  $i$  when reporting defector  $j$ :

$$\begin{aligned} u_i(\pi_i | \pi_j) | \text{report} &= \pi_i | \text{report} - \alpha_i \cdot \max\{\pi_j | \text{reported} - \pi_i | \text{report}, 0\} \\ &= 4 + (\text{Expected payoff in period 2}) - \rho - \alpha_i \cdot [(16 - \rho b_i) - (4 + (\text{Expected payoff in period 2}) - \rho)]. \end{aligned}$$

(ii) The utility of cooperator  $i$  when not reporting defector  $j$ :

$$\begin{aligned} u_i(\pi_i | \pi_j) | \text{not report} &= \pi_i | \text{not report} - \alpha_i \cdot \max\{\pi_j | \text{not reported} - \pi_i | \text{not report}, 0\} \\ &= 4 + (\text{Expected payoff in period 2}) - \alpha_i \cdot [(16 + 6a_i - \rho \cdot b_i) - (4 + (\text{Expected payoff in period 2}))] \end{aligned}$$

Here, cooperator  $i$  decides to report defector  $j$  if  $u_i(\pi_i | \pi_j) | \text{report} > u_i(\pi_i | \pi_j) | \text{not report}$ . In other words,

$$u_i(\pi_i | \pi_j) | \text{report} - u_i(\pi_i | \pi_j) | \text{not report} = -\rho - \alpha_i \cdot [-6a_i - (-\rho)]$$

$$= 6\alpha_i \cdot a_i - (1 + \alpha_i) \cdot \rho > 0.$$

Thus, we have:  $\alpha_i > \frac{\rho}{6a_i - \rho}$ .

Note that this condition is not affected by whether  $i$  “optimistically” assumes that a cooperator selects X when presented with a free 2<sup>nd</sup> choice and meeting another cooperator or “pessimistically” assumes that all players select Y when presented with a free 2<sup>nd</sup> choice.

Reporting Cost ( $\rho$ )	$\alpha_i$	Minimum $a_i$ that makes reporting her YY-choosing counterpart utility maximizing
1.0	0.5	0.500
0.5	0.5	0.250
0.05	0.5	0.025
0.00	0.5	0.000
1.0	1.0	0.333
0.5	1.0	0.167
0.05	1.0	0.017
0.00	0.5	0.000

#### A.4.3. Case 3 – A defector meets a cooperator in period 1

**PROPOSITION A3:** *Suppose that  $\beta_i < 1$ . Then, regardless of which belief (either pessimistic or optimistic) a defector  $i$  forms, the defector never reports her matched cooperator.*

Proof:

Suppose that subject  $i$  chose YY and has been matched with subject  $j$  who selected XX in period 1. Then, defector  $i$  received a payoff of 11 points in that period; cooperator  $j$  received a payoff of 4 points in that period. We examine the conditions under which defector  $i$  reports his matched cooperator  $j$ . We consider the two assumptions on her beliefs.

(a) Suppose that defector  $i$  forms the pessimistic belief. Under this assumption, defector  $i$ 's total expected payoff is calculated as:

$$\begin{aligned}\pi_i(XX) &= 11 - \rho \cdot 1_{report} + (\text{Expected payoff in period 2}) \\ &= 16 - \rho \cdot 1_{report} + 6a_i - 6a_i c_i.\end{aligned}$$

By contrast, defector  $i$ 's belief about cooperator  $j$ 's material payoff if is dependent on whether  $i$  reports  $j$  or not:

(i)  $i$  reports  $j$ :

$$\pi_j|reported = 4 - \rho c_i + (\text{Expected payoff of } j \text{ in period 2 if } i \text{ reports } j)$$

where *Expected payoff of  $j$  in period 2 if  $i$  reports  $j$*

$$\begin{aligned}&= a_i b_i (5) + a_i (1 - b_i) (4) + (1 - a_i) (c_i) (5) + (1 - a_i) (1 - c_i) (4) \\ &= 4 + a_i b_i + c_i - a_i c_i \text{ (See TABLE A.3).}\end{aligned}$$

Thus, we have:

$$\pi_j|reported = 8 - \rho c_i + a_i b_i + c_i - a_i c_i.$$

(ii)  $i$  does not report  $j$ :

$$\pi_j|not\ reported = 4 - \rho c_i + (\text{Expected payoff of } j \text{ in period 2 if } i \text{ does not report } j)$$

where *Expected payoff of  $j$  in period 2 if  $i$  does not report  $j$*

$$\begin{aligned}&= 11(a_i b_i) + 10(a_i)(1 - b_i) + 5c_i(1 - a_i) + 4(1 - a_i)(1 - c_i) \\ &= 4 + 6a_i + a_i b_i + c_i - a_i c_i \text{ (See TABLE A.4).}\end{aligned}$$

Thus, we have:

$$\pi_j|not\ reported = 8 - \rho c_i + 6a_i + a_i b_i + c_i - a_i c_i.$$

First, we find that defector  $i$ 's material payoff is always bigger than cooperator  $j$ 's if  $i$  reports  $j$ . This is because:

$$\pi_i|report - \pi_j|reported = 8 - [\rho(1 - c_i) + a_i(b_i - c_i) + c_i - 6a_i(1 - c_i)] > 0.$$

Second, we also find that  $i$ 's material payoff is always bigger than  $j$ 's if  $i$  does not report  $j$ . This is because:

$$\begin{aligned}\pi_i|not\ report - \pi_j|not\ reported &= 8 - [6a_i c_i - \rho c_i + a_i b_i - a_i c_i + c_i] \\ &= 8 - [6a_i c_i + c_i(1 - \rho) - a_i(c_i - b_i)] > 0.\end{aligned}$$

In other words, regardless of whether  $i$  reports  $j$  or not, defector  $i$ 's material payoff is always bigger than cooperator  $j$ 's.

Defector  $i$ 's Fehr-Schmidt expected utility when  $i$  chooses to report  $j$  is calculated as:

$$\begin{aligned} u(\pi_i|\pi_j) |report &= \pi_i|report - \beta_i \max\{\pi_i|report - \pi_j|reported, 0\} \\ &= 16 - \rho + 6a_i - 6a_i c_i - \beta_i(8 - \rho + 6a_i - 6a_i c_i + \rho c_i - a_i b_i - c_i + a_i c_i). \end{aligned}$$

By contrast,  $i$ 's Fehr-Schmidt expected utility when he chooses not to report  $j$  is calculated as:

$$\begin{aligned} u(\pi_i|\pi_j) |not report &= \pi_i|not report - \beta_i \max\{\pi_i|not report - \pi_j|not reported, 0\} \\ &= 16 + 6a_i - 6a_i c_i - \beta_i(8 - 6a_i c_i + \rho c_i - a_i b_i + a_i c_i - c_i). \end{aligned}$$

Here,  $i$  reports  $j$  if and only if:

$$\begin{aligned} u(\pi_i|\pi_j) |report - u(\pi_i|\pi_j) |not report &> 0, \text{ or} \\ \rho \beta_i &> \rho + 6a_i \beta_i. \end{aligned} \tag{A17}$$

Condition (A17) does not hold (i.e., a defector would not report a cooperator) as we are assuming that  $\beta_i < 1$ .

(b) Suppose that defector  $i$  forms the optimistic belief. Then, defector  $i$ 's total expected payoff is calculated as:

$$\begin{aligned} \pi_i(XX) &= 11 - \rho \cdot 1_{report} + (\text{Expected payoff of } i \text{ in period 2}) \\ &= 16 - \rho \cdot 1_{report} + 6a_i - 6a_i c_i. \end{aligned}$$

Here, the expected payoff of  $i$  in period 2 obtained from TABLE A2.

By contrast,  $i$ 's belief about cooperator  $j$ 's material payoff may be dependent on whether  $i$  reports  $j$  or not.

(i)  $i$  reports  $j$ :

$$\pi_j|reported = 4 - \rho c_i + (\text{Expected payoff of } j \text{ in period 2 if } i \text{ reports } j)$$

where

$$\begin{aligned} \text{Expected payoff of } j \text{ in period 2 if } i \text{ reports } j &= 10a_i + (1 - a_i)(c_i)(5) + (1 - a_i)(1 - c_i)4 \\ &= 4 + 6a_i + c_i - a_i c_i \text{ (see TABLE A.5).} \end{aligned}$$

Thus, we have:

$$\pi_j|reported = 8 - \rho c_i + 6a_i + c_i - a_i c_i.$$

(ii)  $i$  does not report  $j$ :

$$\pi_j|not reported = 4 - \rho c_i + (\text{Expected payoff of } j \text{ in period 2 if } i \text{ does not report } j)$$

where *Expected payoff of  $j$  in period 2 if  $i$  does not report  $j$*

$$\begin{aligned} &= 10a_i + (1 - a_i)(c_i)(5) + (1 - a_i)(1 - c_i)4 \\ &= 4 + 6a_i + c_i - a_i c_i \text{ (See TABLE A.6).} \end{aligned}$$

Thus, we have:

$$\pi_j|reported = 8 - \rho c_i + 6a_i + c_i - a_i c_i.$$

Therefore, we find that the expected payoff of cooperator  $j$  is the same, whether  $i$  reports  $j$  or not. This is because even if  $i$  reports  $j$ , due to optimistic belief, another cooperator will choose the same action if she is matched with  $j$  in period 2.

Defector  $i$ 's material payoff is bigger than  $j$ 's if  $i$  reports  $j$ . This is because:

$$\pi_i|report - \pi_j|reported = 8 - [\rho(1 - c_i) + (1 - a_i)c_i + 6a_i c_i] > 0.$$

Likewise,  $i$ 's material payoff is bigger than  $j$ 's if  $i$  does not report  $j$ . This is because:

$$\pi_i|not\ report - \pi_j|not\ reported = 8 - [(1 - \rho)c_i + 5a_i c_i] > 0.$$

Defector  $i$ 's Fehr-Schmidt expected utility when  $i$  chooses to report  $j$  is calculated as:

$$\begin{aligned} u(\pi_i|\pi_j) |report &= \pi_i|report - \beta_i \max\{\pi_i|report - \pi_j|reported, 0\} \\ &= 16 - \rho + 6a_i - 6a_i c_i - \beta_i(8 - \rho - 5a_i c_i + \rho c_i - c_i). \end{aligned}$$

By contrast, defector  $i$ 's Fehr-Schmidt expected utility when  $i$  chooses not to report  $j$  is calculated as:

$$\begin{aligned} u(\pi_i|\pi_j) |not\ report &= \pi_i|not\ report - \beta_i \max\{\pi_i|not\ report - \pi_j|not\ reported, 0\} \\ &= 16 + 6a_i - 6a_i c_i - \beta_i(16 + 6a_i - 6a_i c_i - 8 + \rho c_i - 6a_i - c_i + a_i c_i) \\ &= 16 + 6a_i - 6a_i c_i - \beta_i(8 - 5a_i c_i + \rho c_i - c_i). \end{aligned}$$

We show that  $i$  never reports  $j$  as we are assuming that  $\beta_i < 1$ . To see this,

$$u(\pi_i|\pi_j) |report - u(\pi_i|\pi_j) |not\ report = -\rho - \beta_i(-\rho) = \rho(\beta_i - 1) < 0.$$

□

#### A.4.4. Case 4 – A defector meets a defector in period 1

PROPOSITION A4: *Suppose that  $\beta_i < 1$ . Then, regardless of which belief, either pessimistic or optimistic, a defector  $i$  forms, a defector reports (does not report) her matched defector if and only if the following two conditions hold (do not hold):*

$$6a_i > \rho \text{ and } -\rho + ((\alpha_i + \beta_i)c_i - \beta_i)(6a_i - \rho) > 0.$$

*A comparison between the conditions in Propositions A1 and Proposition A4 suggests that reporting will be more frequent in Case 4 if  $c_i > b_i$*

Proof:

Suppose that subject  $i$  has been matched with subject  $j$  in period 1 and both subjects select YY. Then, subjects  $i$  and  $j$  each receive a payoff of 5 points in that period. We examine the conditions under which defector  $i$  reports his matched defector  $j$ .

Defector  $i$ 's total expected payoff is calculated as:

$$\begin{aligned}\pi_i(YY) &= 5 - \rho \cdot 1_{\text{report}} + (\text{Expected payoff of } i \text{ in period 2}) \\ &= 10 - \rho \cdot 1_{\text{report}} + 6a_i - 6a_i c_i.\end{aligned}$$

Defector  $i$ 's belief about  $j$ 's material payoff is dependent on whether  $i$  reports  $j$  or not.

(i) If  $i$  reports  $j$ :

$$\pi_j|\text{reported} = 5 - \rho c_i + (\text{Expected payoff of } j \text{ in period 2 if } i \text{ reports } j)$$

where *Expected payoff of  $j$  in period 2 if  $i$  reports  $j$*  = 5.

Thus, we have:

$$\pi_j|\text{reported} = 10 - \rho c_i.$$

(ii) If  $i$  does not report  $j$ :

$$\pi_j|\text{not reported} = 5 - \rho c_i + (\text{Expected payoff of } j \text{ in period 2 if } i \text{ does not report } j)$$

where

$$\text{Expected payoff of } j \text{ in period 2 if } i \text{ does not report } j = 11a_i + (1 - a_i)5.$$

Thus, we have:

$$\pi_j|\text{not reported} = 10 + 6a_i - \rho c_i.$$

We claim that if  $i$  reports  $j$ ,  $i$ 's material payoff is bigger than  $j$ 's if  $6a_i > \rho$ . To see this,

$$\pi_i|\text{report} - \pi_j|\text{reported} = [-\rho + 6a_i(1 - c_i)] - (-\rho c_i) = (6a_i - \rho)(1 - c_i) > 0 \text{ if and only if } 6a_i > \rho.$$

If  $i$  does not report  $j$ ,  $i$ 's material payoff is bigger than  $j$ 's if  $\rho > 6a_i$ . This is because:

$$\pi_i|\text{not report} - \pi_j|\text{not reported} = \rho - 6a_i > 0 \text{ if and only if } \rho > 6a_i.$$

Thus, we need to consider the two cases:  $6a_i > \rho$  and  $6a_i < \rho$ .



Suppose that  $\rho > 6a_i$ . Then, defector  $i$ 's Fehr-Schmidt expected utility when  $i$  chooses to report  $j$  is calculated as:

$$\begin{aligned} u(\pi_i|\pi_j) |report &= \pi_i|report - \alpha_i \max\{\pi_j|reported - \pi_i|report, 0\} \\ &= 10 - \rho + 6a_i - 6a_i c_i - \alpha_i(-\rho c_i + \rho - 6a_i + 6a_i c_i). \end{aligned}$$

By contrast,  $i$ 's Fehr-Schmidt expected utility when  $i$  chooses not to report is calculated as:

$$\begin{aligned} u(\pi_i|\pi_j) |not report &= \pi_i|not report - \beta_i \max\{\pi_i|not report - \pi_j|not reported, 0\} \\ &= 10 + 6a_i - 6a_i c_i - \beta_i(-\rho + 6a_i - 6a_i c_i + \rho c_i). \end{aligned}$$

Therefore,  $i$  reports  $j$  if and only if:

$$\begin{aligned} u(\pi_i|\pi_j) |report - u(\pi_i|\pi_j)|not report &> 0, \text{ or} \\ -\rho - \alpha_i(-\rho c_i + \rho - 6a_i + 6a_i c_i) &> -\beta_i(-\rho + 6a_i - 6a_i c_i + \rho c_i), \end{aligned}$$

a condition that never holds as this condition can be re-arranged as:

$$-\rho > (\alpha_i + \beta_i)(1 - c_i)(\rho - 6a_i),$$

but the right-hand side is non-negative. In other words,  $i$  does not report  $j$  if  $\rho > 6a_i$ .

Suppose instead that  $\rho < 6a_i$ . Then, defector  $i$ 's Fehr-Schmidt expected utility when  $i$  chooses to report  $j$  is calculated as:

$$\begin{aligned} u(\pi_i|\pi_j) |report &= \pi_i|report - \beta_i \max\{\pi_i|report - \pi_j|reported, 0\} \\ &= 10 - \rho + 6a_i - 6a_i c_i - \beta_i(-\rho + 6a_i - 6a_i c_i + \rho c_i). \end{aligned}$$

By contrast,  $i$ 's Fehr-Schmidt expected utility when  $i$  chooses not to report  $j$  is calculated as:

$$\begin{aligned} u(\pi_i|\pi_j) |not report &= \pi_i|not report - \alpha_i \max\{\pi_j|not reported - \pi_i|not report, 0\} \\ &= 10 + 6a_i - 6a_i c_i - \alpha_i(-\rho c_i + 6a_i c_i). \end{aligned}$$

Thus, we find:  $i$  reports  $j$  if and only if:

$$\begin{aligned} u(\pi_i|\pi_j) |report - u(\pi_i|\pi_j)|not report &> 0, \text{ or} \\ -\rho - \beta_i(-\rho + 6a_i - 6a_i c_i + \rho c_i) &> -\alpha_i(-\rho c_i + 6a_i c_i), \text{ or} \\ -\rho - \beta_i(1 - c_i)(-\rho + 6a_i) &> -\alpha_i c_i(-\rho + 6a_i), \text{ or} \\ -\rho + (-\rho + 6a_i)((\alpha_i + \beta_i)c_i - \beta_i) &> 0. \end{aligned}$$

This means that only when  $c_i$  is sufficiently large and also when  $a_i$  is large enough that  $6a_i > \rho$ , a defector reports her matched defector.

□

COROLLARY A4: Suppose that  $\beta_i < 1$ . Then, subject  $i$  engages in more costly reporting in Case 4 than in Case 1 if  $b_i < c_i$  (subject  $i$  engages in more costly reporting in Case 1 than in Case 4 if  $b_i > c_i$ ).

Proof:

From Proposition A4, an additional requirement that defector  $i$  reports another defector  $j$  is that:  $c_i > \frac{\beta_i}{\alpha_i + \beta_i}$ ; otherwise,  $-\rho + ((\alpha_i + \beta_i)c_i - \beta_i)(6a_i - \rho) \leq 0$ .

The condition:  $-\rho + ((\alpha_i + \beta_i)c_i - \beta_i)(6a_i - \rho) > 0$  in Proposition A4 can be re-arranged to:

$$a_i > \frac{\rho}{6} + \frac{\rho}{6[(\alpha_i + \beta_i)c_i - \beta_i]}. \quad (\text{A18})$$

The right hand side of condition (A18) is less (greater) than  $\frac{\rho}{6} + \frac{\rho}{6[(\alpha_i + \beta_i)b_i - \beta_i]}$  when  $b_i < c_i$  ( $b_i > c_i$ ). This means that condition (A18) is weaker (stronger) than the condition:  $-\rho + ((\alpha_i + \beta_i)b_i - \beta_i)(6a_i - \rho) > 0$  in Proposition A1. This means, we observe more (less) costly reporting in Case 4 than in Case 1 if  $b_i < c_i$  ( $b_i > c_i$ ).

## Appendix B: Additional Tables and Figures

**TABLE B.1:** Summary of the subjects' initial choices (XX vs. YY) and reporting decisions

### (A) Decision Data

	Treatment by Reporting Cost				
	HC	MC	LC	Subtotal (Costly reporting)	NC
Reporting Cost	\$1.00	\$0.50	\$0.05	----	\$0.00
(i) Number of Subjects	38 (100%)	38 (100%)	40 (100%)	116 (100%)	36 (100%)
(ii) Number of Cooperators (percentage)	23 (60.5%)	19 (50.0%)	21 (52.5%)	63 (54.3%)	23 (63.9%)
<i>Number of Cooperators being reported [percentage]</i>	0 [0.0%]	1 [5.3%]	2 [9.5%]	3 [4.8%]	12 [52.2%]
<i>Number of Cooperators that report [percentage]</i>	6 [26.1%]	7 [36.8%]	7 [33.3%]	20 [31.7%]	16 [69.6%]
• <i>Number of Cooperators that face Cooperators [percentage]</i>	14 [60.9%]	10 [52.6%]	10 [47.6%]	34 [54.0%]	14 [60.9%]
<i>Number of Cooperators that report Cooperators (percentage of reporting)</i>	0 (0.0%)	1 (10.0%)	2 (20.0%)	3 (8.8%)	8 (57.1%)
• <i>Number of Cooperators that face Defectors [percentage]</i>	9 [39.1%]	9 [47.4%]	11 [52.4%]	29 [46.0%]	9 [39.1%]
<i>Number of Cooperators that report Defectors (percentage of reporting)</i>	6 (66.7%)	6 (66.7%)	5 (45.5%)	17 (58.6%)	8 (88.9%)
(iii) Number of Defectors (percentage)	15 (39.5%)	19 (50.0%)	19 (47.5%)	53 (45.7%)	13 (36.1%)
<i>Number of Defectors being reported [percentage]</i>	7 [46.7%]	8 [42.1%]	6 [31.6%]	21 [39.6%]	11 [84.6%]
<i>Number of Defectors that report [percentage]</i>	1 [6.67%]	2 [10.5%]	1 [5.26%]	4 [7.5%]	7 [53.9%]
• <i>Number of Defectors that face Cooperators [percentage]</i>	9 [60.0%]	9 [47.4%]	11 [57.9%]	29 [54.7%]	9 [69.2%]
<i>Number of Defectors that report Cooperators (percentage of reporting)</i>	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	4 (44.4%)
• <i>Number of Defectors that face Defectors [percentage]</i>	6 [40.0%]	10 [52.6%]	8 [42.1%]	24 [45.3%]	4 [30.8%]
<i>Number of Defectors that report Defectors (percentage of reporting)</i>	1 (16.7%)	2 (20.0%)	1 (12.5%)	4 (16.7%)	3 (72.5%)

(B) Tests of equality of percentages of cooperators or defectors being reported, across treatments

B.1. Costly reporting treatments versus no cost (NC) reporting treatment

	HC vs. NC treatment	MC vs. NC treatment	LC vs. NC treatment	The three costly reporting treatments vs. NC treatment
(i) The % of Cooperators being reported	.0001***	.0011***	.0024***	.0000***
• <i>The % of Cooperators being reported by XX choosers</i>	.0008***	.0187**	.0688*	.0003***
• <i>The % of Cooperators being reported by YY choosers</i>	.0233**	.0233**	.0134**	.0001***
(ii) The % of Defectors being reported	.0366**	.0162**	.0031***	.0036***
• <i>The % of Defectors being reported by XX choosers</i>	.2568	.2568	.0428**	.0945*
• <i>The % of Defectors being reported by YY choosers</i>	.0651*	.0524*	.0304**	.0126**

Notes: Panel (B) reports two-sample test of proportion results. The numbers are  $p$ -values (two-sided).

\*\*\* Significant at the 1 percent level.

\*\* Significant at the 5 percent level.

\* Significant at the 10 percent level.

## B.2. Comparison of reporting frequency by defectors vs. by cooperators, within treatments

	HC treatment	MC treatment	LC treatment	Three costly reporting treatments	NC treatment
• The % of defectors being reported by XX choosers versus the % of defectors being reported by YY choosers	.0572*	.0397**	.1271	.0019***	.5218
• The % of defectors being reported by XX choosers versus the % of cooperators being reported by XX choosers	.0004***	.0106**	.2165	.0000***	.1063
• The % of defectors being reported by XX choosers versus the % of cooperators being reported by YY choosers	.0027***	.0027***	.0110**	.0000***	.0455**
• The % of defectors being reported by YY choosers versus the % of cooperators being reported by XX choosers	.1171	.5312	.5312	.3665	.5182
• The % of defectors being reported by YY choosers versus the % of cooperators being reported by YY choosers	.2049	.1561	.2283	.0222**	.3077
• The % of cooperators being reported by XX choosers versus the % of cooperators being reported by YY choosers	n.a. <sup>#1</sup>	.3297	.1189	.1012	.5518

Notes: Panel (B) reports two-sample test of proportion results. The numbers are  $p$ -values (two-sided). <sup>#1</sup> All subjects in these case did not report their counterparts' decisions.

\*\*\* Significant at the 1 percent level.

\*\* Significant at the 5 percent level.

\* Significant at the 10 percent level.

## (C) Comparison of cooperation frequency across the treatments

		Treatment			
		HC	MC	LC	NC
Treatment	HC	----	.3561	.4749	.7656
	MC	----	----	.8253	.2281
	LC	----	----	----	.3153
	Subtotal (HC,MC,LC)	----	----	----	3111

Notes: Panel (C) reports two-sample test of proportion results. The numbers are  $p$ -values (two-sided).

\*\*\* Significant at the 1 percent level.

\*\* Significant at the 5 percent level.

\* Significant at the 10 percent level.

**TABLE B.2:** Tests of Prediction H-SRC

(a) The fraction of cooperators (H-SRC predicts no cooperation)

	Costly Reporting Treatments				NC
	HC	MC	LC	Subtotal	
<i>p</i> -value (two-sided) <sup>1</sup>	.000***	.000***	.000***	.000***	.000***

*Notes:* Binomial probability test results. We adopt the conservative null hypothesis that the probability of choosing XX equals 5%, assuming that errors occur with a probability of 5%. \*, \*\*, and \*\*\* indicate significance at the .10 level, at the .05 level and at the .01 level, respectively. The hypothesis that not more than 5% cooperate is rejected at the .01 level in all treatments.

(b) The fraction of cooperators that report XX-choosing counterparts (H-SRC predicts no reporting in HC, MC and LC)

	Costly Reporting Treatments				NC
	HC	MC	LC	Subtotal	
<i>p</i> -value (two-sided) <sup>1</sup>	1.000	.401	.086*	.241	.000***

*Notes:* Binomial probability test results. We adopt the conservative null hypothesis that the probability that cooperators report the initial choices of their XX-choosing partners equals 5%, assuming that errors occur with a probability of 5%. \*, \*\*, and \*\*\* indicate significance at the .10 level, at the .05 level and at the .01 level, respectively. The hypothesis of no reporting apart from error is not rejected in HC and MC and in the costly reporting treatments when pooled, and is rejected at the 10% level only in LC treatment.

(c) The fraction of cooperators that report YY-choosing counterparts (H-SRC predicts no reporting in HC, MC and LC)

	Costly Reporting Treatments				NC
	HC	MC	LC	Subtotal	
<i>p</i> -value (two-sided) <sup>1</sup>	.000***	.000***	.000***	.000***	.000***

*Notes:* Binomial probability test results. We adopt the conservative null hypothesis that the probability that cooperators report the initial choices of their YY-choosing partners equals 5%, assuming that errors occur with a probability of 5%. \*, \*\*, and \*\*\* indicate significance at the .10 level, at the .05 level and at the .01 level, respectively. The prediction of no reporting apart from errors is rejected at the .01 level in all cases.

(d) The fraction of defectors that report XX-choosing counterparts (H-SRC predicts no reporting in HC, MC and LC)

	Costly Reporting Treatments				NC
	HC	MC	LC	Subtotal	
<i>p</i> -value (two-sided) <sup>1</sup>	1.000	1.000	1.000	.401	.001***

*Notes:* Binomial probability test results. We adopt the conservative null hypothesis that the probability that defectors report the initial choices of their XX-choosing partners equals 5%, assuming that errors occur with a probability of 5%. \*, \*\*, and \*\*\* indicate significance at the .10 level, at the .05 level and at the .01 level, respectively. The hypothesis of no reporting apart from errors is not rejected in the costly reporting treatments, but a hypothesis of no reporting in NC (which is not part of H-SRC as such) would be rejected in the case of NC treatment.

(e) The fraction of defectors that report YY-choosing counterparts (H-SRC predicts no reporting in HC, MC and LC)

	Costly Reporting Treatments				NC
	HC	MC	LC	Subtotal	
<i>p</i> -value (two-sided) <sup>1</sup>	.265	.086*	.337	.0298**	.000***

*Notes:* Binomial probability test results. We adopt the conservative null hypothesis that the probability that defectors report the initial choices of their YY-choosing partners equals 5%, assuming that errors occur with a probability of 5%. \*, \*\*, and \*\*\* indicate significance at the .10 level, at the .05 level and at the .01 level, respectively. The hypothesis that there is no reporting except for errors is rejected at the 5% level for the pooled costly reporting treatments, but at the 10% level only for the MC treatment taken alone. A hypothesis of no costly reporting (not part of H-SRC as such) would be rejected in the case of NC treatment.

**TABLE B.3:** Cooperation decisions & predictions assuming payoff maximization and self-reported beliefs

	Treatment by Reporting Cost					
	HC	MC	LC	Subtotal	NC	Total
<b>Number of XX-choosers</b>	23	19	21	63	23	86
(i) Number of subjects for whom LHS of (1) > 2 <sup>a</sup>	6 (15.8%)	8 (21.1%)	13 (32.5%)	27 (23.3%)	11 (30.6%)	38 (25.0%)
(i-1) Number of subjects who chose XX:	6	6	13	25	10	35
<i>Average belief <math>a_i</math> for those in (i-1) – #1</i>	82.2%	84.5%	78.5%	80.8%	80.5%	80.7%
(i-2) Subjects who select XX and have LHS of (1) > 2 <sup>a</sup> as share of all subjects who select XX	26.1%	31.6%	61.9%	39.7%	43.5%	40.7%
(ii) Subjects who select XX and have LHS of (1') > 2 <sup>b</sup> but not LHS of (1) > 2	4 (17.4%)	6 (31.6%)	6 (28.6%)	16 (25.4%)	8 (34.8%)	24 (27.9%)
(iii) Subjects not in (i-1) or (ii) for whom FS model predicts cooperation (based on condition (4)) if $\alpha = 1$ , $\beta = 0.6$	4 (17.4%)	4 (21.1%)	1 (4.8%)	9 (14.3%)	1 (4.3%)	10 (11.6%)
(iv) Subjects not in (i-1), (ii) or (iii) for whom FS model predicts cooperation (based on condition (4)) if $\alpha = 1$ , $\beta = 1$	2 (8.7%)	0 (0.0%)	1 (4.8%)	3 (4.8%)	1 (4.3%)	4 (4.7%)
(v) Same as (iv) except assuming $\alpha = 0$ , $\beta = 1$	6 (26.1%)	3 (15.8%)	0 (0.0%)	9 (14.3%)	3 (13.0%)	12 (14.0%)
(vi) Subject who select XX but not in (i-1) to (v)	1 (4.3%)	0 (0.0%)	0 (0.0%)	1 (1.6%)	0 (0.0%)	1 (1.2%)
(vii) Number of subjects in (i) who chose YY:	0 [0.0%]	2 [25.0%]	0 [0.0%]	2 [7.4%]	1 [9.1%]	3 [7.9%]
<i>Average belief <math>a_i</math> for those in (i-2)</i>	<i>n.a.</i>	55.0%	<i>n.a.</i>	55.0%	60.0%	56.7%
<b>Number of YY-choosers</b>	15	19	19	53	13	66
(viii) Number of subjects for whom LHS of (1) < 2	32	30	26	88	25	113
(viii-1) Number of subjects who chose XX:	17 [53.1%]	13 [43.3%]	8 [30.8%]	38 [43.2%]	13 [52.0%]	51 [45.1%]
<i>Average belief <math>a_i</math> for those in (viii-1) – #2</i>	59.9%	69.2%	76.5%	66.6%	66.2%	66.5%
(viii-2) Number of subjects who chose YY:	15	17	18 <sup>c</sup>	50	12	62
<i>Average belief <math>a_i</math> for those in (viii-2) – #3</i>	41.5%	38.8%	27.7%	35.6%	23.5%	33.2%
(viii-3) Subjects selecting YY and having LHS of (1) < 2 as share of all subjects who choose YY	100%	89.5%	94.7%	94.5%	92.3%	94.1%
Two sample test of proportion <sup>1</sup>						
H <sub>0</sub> : % of those who chose YY among (i) = % of those who chose XX among (viii)	.0163**	.3459	.0282**	.001***	.0150**	0.000***
Mann-Whitney tests <sup>1</sup>						
H <sub>0</sub> : term (#1) = term (#2)	.0676*	.0933*	.7977	.0141**	.0861*	.0033***
H <sub>0</sub> : term (#2) = term (#3)	.0850*	.0010***	.0005***	.0000***	.0003***	.0000***

Notes: <sup>1</sup> The numbers are *p*-values (two-sided).

<sup>a</sup> LHS of (1) > 2 is the condition for cooperation to be payoff maximizing given the individual's self-reported beliefs and assuming those free to choose select defect for second interaction.

<sup>b</sup> LHS of (1') > 2 is the condition for cooperation to be payoff maximizing given the individual's self-reported beliefs and assuming that initial cooperators (XX-choosers) free to revise their decision choose to cooperate in second interaction if get report that counterpart cooperated.



<sup>c</sup> There was one subject whose LHS of (1) is equal to 2 (with  $a = 20$ ,  $b = 0$  and  $c = 100$ ) and who selected YY.

\*\*\* Significant at the 1 percent level.

\*\* Significant at the 5 percent level.

\* Significant at the 10 percent level.

**TABLE B.4:** Average beliefs ( $a_i$ ,  $b_i$  and  $c_i$ ) and tests for differences, with cooperators and defectors distinguished

(i) Average Beliefs

	Costly Reporting Treatments				NC
	HC	MC	LC	Subtotal (HC,MC,LC)	
(i.1) Cooperators					
[1] Average $a_i$	65.7%	74.0%	77.7%	72.2%	72.4%
[2] Average $b_i$	21.0%	27.9%	29.7%	26.0%	39.3%
[3] Average $c_i$	39.8%	52.4%	75.6%	55.5%	75.1%
(i.2) Defectors					
[4] Average $a_i$	41.5%	40.5%	27.3%	36.0%	26.3%
[5] Average $b_i$	32.5%	28.4%	27.2%	29.1%	49.9%
[6] Average $c_i$	26.1%	35.8%	25.4%	29.3%	67.9%

(ii) Test Results<sup>1</sup>

We tested the difference in the beliefs of cooperators vs. those of defectors, by treatment. A significant result means beliefs of cooperators differ significantly from those of defectors.

	HC	MC	LC	Subtotal (HC,MC,LC)	NC
[1] = [4]	.0174**	.0000***	.0000***	.0000***	.0000***
[2] = [5]	.1298	.5650	.6930	.4575	.2817
[3] = [6]	.2805	.1610	.0000***	.0001***	.2184

Notes: <sup>1</sup> Mann-Whitney test results. The numbers are  $p$ -values (two-sided). The test results comparing beliefs across the treatments are found in Appendix TABLE B.5.

\*\*\* Significant at the 1 percent level.

\*\* Significant at the 5 percent level.

\* Significant at the 10 percent level.

**TABLE B.5:** Comparison of the subjects' beliefs across treatments by initial choice (XX or YY)  
(Supplementing TABLE B.4 of this Appendix)

(1) Belief  $a_i$

We tested the differences in the average belief  $a$  between the treatments, for cooperators (panel 1a) and for defectors (panel 1b).

(1a) Cooperators

		Treatment			
		HC	MC	LC	NC
Treatment	HC	----	.3480	.1569	.4276
	MC	----	----	.7639	.7320
	LC	----	----	----	.4778
	Subtotal (HC,MC,LC)	----	----	----	.9415
	NC	----	----	----	----

(1b) Defectors

		Treatment			
		HC	MC	LC	NC
Treatment	HC	----	.9445	.1386	.1959
	MC	----	----	.0514*	.0766*
	LC	----	----	----	.8024
	Subtotal (HC,MC,LC)	----	----	----	.2752
	NC	----	----	----	----

*Notes.* Two-sided individual-level Mann-Whitney tests. Numbers in panels are  $p$ -values. The insignificant results for most cases means that beliefs of cooperators (defectors) about the % that would cooperate do not tend to differ across treatments.

\*\*\* Significant at the 1 percent level.

\*\* Significant at the 5 percent level.

\* Significant at the 10 percent level.

(2) Belief  $b_i$

We tested the differences in the average belief  $b$  between the treatments for cooperators (panel 2a) and defectors (panel 2b).

(2a) Cooperators

		Treatment			
		HC	MC	LC	NC
Treatment	HC	----	.6938	.4504	.0317**
	MC	----	----	.7958	.1543
	LC	----	----	----	.2830
	Subtotal (HC,MC,LC)	----	----	----	.0513*

(2b) Defectors

		Treatment			
		HC	MC	LC	NC
Treatment	HC	----	.6745	.3294	.0970*
	MC	----	----	.5672	.0307**
	LC	----	----	----	.0269**
	Subtotal (HC,MC,LC)	----	----	----	.0158**

*Notes.* Two-sided individual-level Mann-Whitney tests. Numbers in the panels are  $p$ -values. Results indicate that defectors especially had higher expectations of share of cooperators that would be reported in the no cost than in the various costly reporting treatments.

\*\*\* Significant at the 1 percent level.

\*\* Significant at the 5 percent level.

\* Significant at the 10 percent level.

### (3) Belief $c_i$

We tested the differences in the average belief  $c$  between the treatments for each of the cooperators and the YY choosers.

#### (3a) For cooperators

		Treatment			
		HC	MC	LC	NC
Treatment	HC	----	.2492	.0005***	.0004***
	MC	----	----	.0687*	.0334**
	LC	----	----	----	.7765
	Subtotal (HC,MC,LC)	----	----	----	.0118**

#### (3a) For defectors

		Treatment			
		HC	MC	LC	NC
Treatment	HC	----	.2645	.4634	.0006***
	MC	----	----	.1420	.0037***
	LC	----	----	----	.0010***
	Subtotal (HC,MC,LC)	----	----	----	.0001***

*Notes.* Two-sided, individual-level Mann-Whitney tests. Numbers in the panels are  $p$ -values.

\*\*\* Significant at the 1 percent level.

\*\* Significant at the 5 percent level.

\* Significant at the 10 percent level.

**TABLE B.6:** The Subjects' Initial Choices of XX or YY and their Three Kinds of Beliefs  
(Supplementing Panel (ii) of TABLE B.4 of this Appendix)

In this table, we conducted a regression analysis to explore the relationship between the subjects' initial choices and their beliefs, instead of performing non-parametric tests as shown in Panel (ii) of TABLE B.5.

Dependent Variable: A dummy that equals 1 if a subject choose to play XX; 0 otherwise

Variables	(1)	(2)	(3)	(4)
Belief $a_i$ (Belief of subject $i$ about the fraction of Cooperators in his or her session)	.011*** (.0011)	----	----	.010*** (.0011)
Belief $b_i$ (Belief of subject $i$ about the fraction of Cooperators being reported)	----	-.0012 (.0015)	----	-.0013 (.0011)
Belief $c_i$ (Belief of subject $i$ about the fraction of Defectors being reported)	----	----	.0049*** (.0011)	.0020** (.00095)
Constant	-0.054*** (.066)	.60*** (.061)	.32*** (.067)	-.066 (.077)
# of Observation	152	152	152	152
F	110.73	.65	19.91	39.51
Prob > F	.0000	..4202	.0000	.0000
Adjusted R-squared	.4209	-.0023	.1113	.4334

Notes: Linear regressions. Observations of all reporting treatments (HC, MC, LC and NC) are included in the regressions.

\*\*\* Significant at the 1 percent level.

\*\* Significant at the 5 percent level.

\* Significant at the 10 percent level.

**TABLE B.7:** The Deviation of Prediction H-SPO (Supplementing Appendix TABLE B.3 of this manuscript)

	Treatment by Reporting Cost					
	Costly Reporting Treatments				NC	Total
	HC	MC	LC	Subtotal		
(i) Number of subjects whose $E[\pi_i(XX)] - E[\pi_i(YY)] > 0$	6	8	13	27	11	38
Number of subjects who chose YY:	0 [0.0%]	2 [25.0%]	0 [0.0%]	2 [7.4%]	1 [9.1%]	3 [7.9%]
$p$ -value (two-sided) for binomial probability tests to the null that a subject commits this error with a probability of 5% (i.e., s/he chooses YY with a 5% probability even if $E[\pi_i(XX)] - E[\pi_i(YY)] > 0$ ).	1.000	.057*	1.000	.394	.431	.438
(ii) Number of subjects whose $E[\pi_i(YY)] - E[\pi_i(XX)] > 0$	32	30	26	88	25	113
Number of subjects who chose XX:	17 [53.1%]	13 [43.3%]	8 [30.8%]	38 [43.2%]	13 [52.0%]	51 [45.1%]
$p$ -value (two-sided) for binomial probability tests to the null that a subject commits this error with a probability of 5% (i.e., s/he chooses XX with a 5% probability even if $E[\pi_i(YY)] - E[\pi_i(XX)] > 0$ ).	.000***	.000***	.000***	.000***	.000***	.000***

Notes: \*\*\* Significant at the 1 percent level. \*\* Significant at the 5 percent level. \* Significant at the 10 percent level.

**TABLE B.8:** The Determinants of Reporting Decisions in the Costly Reporting Treatments

Independent Variable	Case 1: Cooperators face XX choosers (1)	Case 2: Cooperators face YY choosers (2)	Case 4: Defectors face YY choosers (3)
Belief $a_i$ [%]	.0014 (.0016)	.0010 (.0051)	-.0037 (.0041)
Belief $b_i$ [%]	.0087*** (.0013)	-.00047 (.0035)	-.0039 (.0039)
Belief $c_i$ [%]	.00015 (.0011)	.0084*** (.0029)	.013** (.0047)
HC treatment dummy {= 1 for HC treatment; 0 otherwise}	-.027 (.092)	.38 (.22)	-.26 (.22)
MC treatment dummy {= 1 for MC treatment; 0 otherwise}	-.037 (.087)	.38* (.21)	-.21 (.20)
Constant	-.20 (.15)	-.21 (.50)	.23 (.17)
# of Observations	34	29	24
The number of reporting events	3	17	4
F	10.72	2.16	1.82
Prob > F	.0000	.0943	.1590
Adjusted R-Squared	.5955	.1715	.1518

Notes: Linear regressions. There were no reporting event for Case 3 (Defectors face XX choosers) in the HC, MC and LC treatments.

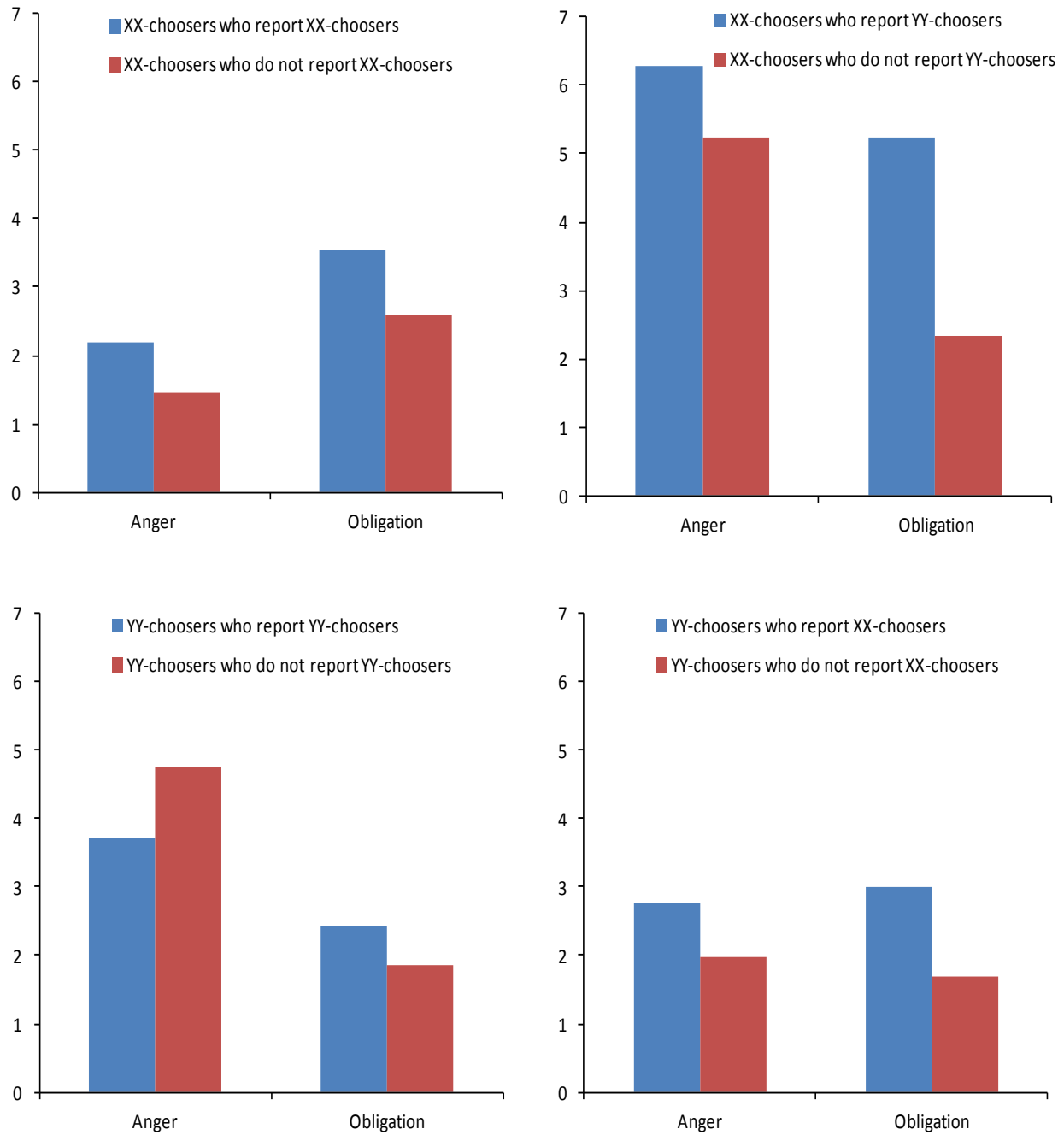
\*\*\* Significant at the 1 percent level.

\*\* Significant at the 5 percent level.

\* Significant at the 10 percent level.



**Fig. B.1:** Average Anger Level and the Feeling of Obligation to Help a Third Person



*Notes:* The Anger variable is a subject's response to the following question: "How did you feel about your first counterpart's decision? Please rate on a scale from 1 = very pleased to 7 = very angry." The Obligation variable is a subject's response to the following question: "Did you feel a sense of obligation to help your first counterpart's next counterpart by sending a report? Please rate on a scale from 1 = did not feel obligated at all, to 7 = felt strongly obligated."

## **Appendix C: Instructions for the LC treatment**

### **Instructions**

You are participating in a decision-making experiment in which you will earn an amount of money that depends on your decisions and on the decisions of other participants.

Please switch off your cell phone. During the experiment, you are not allowed to communicate with other participants.

In the experiment, you will be engaging in two interactions, each with a different, randomly selected, counterpart. Each interaction has the same basic structure, including the amounts of money at stake. Your decisions are anonymous. You will not be told the identities of either of the participants with whom you interact in the experiment, nor will those with whom you interact know your identity. Your decisions will be recorded without any identifiers, and thus, the experimenters also cannot match your decisions with your name. We can assure you that your payoffs will be based only on your own decisions and on the decisions of other actual participants in today's experiment, and that neither a computer program nor members of the experiment team will ever be substituted for other participants. Details will follow after we first explain the nature of the interaction.

#### *Basic Feature of Interactions*

In each of two interactions, each of the two participants who are paired for it decides between two alternative decisions, called X and Y. The amount of money that you will earn from the interaction depends only on your choice and on the choice of the person you are paired with. There are four possibilities:

- (a) If you choose X and your counterpart also chooses X, you earn \$10.
- (b) If you choose Y and your counterpart also chooses Y, you earn \$5.
- (c) If you choose X and your counterpart chooses Y, you earn \$4.
- (d) If you choose Y and your counterpart chooses X, you earn \$11.

(Your counterpart has the same earning formula as you.)

The second interaction is exactly like the first one. In other words, you will be engaging in the identical type of interaction twice, each time with a different counterpart.

A more schematic way of visualizing the possible choices of your counterpart and yourself and the payoffs that would result under each possible set of choices is shown in the table on the last page of these instructions. Please review that table and the instructions so far, and raise your hand if you have a question before we go on with the instructions.

### *Your first decision: choosing XX or YY*

You will be asked to make your choices, either X or Y, for both your first and your second interaction at the outset. Even though the two interactions are separate and are conducted with different counterparts, you are required to make the identical choice for both of them. In other words, you can decide to choose X in both interactions, which we call “XX”, or you can decide to choose Y in both interactions, which we call “YY.” You cannot select “X, then Y” or “Y, then X.”

### *Your second decision: choosing whether to report your counterpart’s action*

Once you have made your initial choice of either XX or YY, you will be randomly paired with a first counterpart and your earnings from your first interaction will be calculated. (The pairing is completely random and cannot be influenced by either your own or your counterpart’s initial choice.) The computer will then inform you of the outcome of your interaction with this first randomly assigned counterpart. The screen in question will remind you of your initial decision of XX or YY, will tell you about your counterpart’s initial choice of XX or YY, and will indicate your and your counterpart’s earnings from the first interaction based on those two sets of decisions.

Following this, and before moving on to the second interaction, you will be asked to decide whether you wish to spend five cents (\$0.05) of your earnings to reveal your first interaction counterpart’s choice **to that person’s next interaction counterpart**. If you choose to report your first counterpart’s action, then that individual’s choice of XX or YY plus his or her earnings in her interaction with you will be made known to his or her partner in the second interaction, potentially affecting their earnings in that interaction as explained next.

Once each participant submits their decision to report or not report the action of their first counterpart, the computer will randomly match you with another participant with whom you’ll engage in Phase 2. (There is zero chance that your second counterpart is the same individual with whom you interacted the first time.)

Please consider the instructions so far, and raise your hand if you have a question.

## **Second Interaction**

### *The no report case:*

Your interaction with your second counterpart will proceed exactly like that with your first counterpart unless at least one of the two of you, yourself or your new counterpart, has been sent a report by the first participant you or he/she interacted with. Specifically, if neither of you are sent a report, neither you nor your counterpart has a new decision to make. The computer will

simply take your initial choice of XX or YY and your counterpart's initial choice of XX or YY, will calculate the appropriate payoffs for that pair of choices, and will credit you with those earnings.

Example: You selected XX, your first counterpart selected YY, your second counterpart selected XX, and neither you nor your second counterpart receives a report about first interaction behavior by one another's initial counterparts. Therefore, neither you nor your second counterpart can alter your decisions. You earn \$4 and your first counterpart earns \$11 from your first interaction. You earn \$10 and your second counterpart earns \$10 from your second interaction. Your total earnings are  $\$4 + \$10 = \$14$ , which, together with the participation fee of \$5, gives you earnings of \$19. Your total earnings are \$18.95 if you chose to report your counterpart's action and \$19.00 if you chose not to report it.

#### *The case of reporting:*

To restate, new choices are possible in your second interaction only if at least one of the two participants in question, yourself or your new counterpart, has received a report about the other's initial action.

A participant who receives a report has a new decision to make.

If you receive a report about your second counterpart's initial action thanks to the decision of their original counterpart, you will be told

- whether your new counterpart had chosen XX or YY
- what your new counterpart earned in his or her first interaction
- whether your new counterpart is in a position to make a fresh decision (like you are) or is unable to make a fresh decision (i.e., is not in receipt of a report from your own first counterpart).

Regardless of whether only you, or both you and your new counterpart, are able to make a fresh decision, your next step is to decide whether to keep to your original choice of X or Y, or to change your choice (from X to Y or from Y to X). If only you can change your choice, then you know your counterpart's decision will be the one he or she made originally, so you can determine the consequence of whatever choice you make with certainty. If both you and your counterpart can change your choices, you will be deciding what to do knowing that your counterpart is simultaneously making a decision and has information about what your initial decision was.

Example 1: You receive the report that your new counterpart chose YY, earned \$11 in his or her first interaction, and is unable to make a new decision. You can choose X for the second interaction, in which case you earn \$4 and your new counterpart earns \$11, or you can choose Y for the second interaction, in which case both you and your new counterpart earn \$5. You can

choose either X or Y in this second interaction, regardless of whether your original choice was XX or YY.

Example 2: Your choice in your first interaction was XX and you do not receive a report. Your counterpart's choice was also XX but your counterpart receives a report about your initial choice. You will be told that you have no decision to make and must wait while your counterpart and (perhaps) others make their choices. If your counterpart selects X for the second interaction, you both earn \$10. If your counterpart selects Y, you earn \$4 and he or she earns \$11.

Example 3: Both you and your counterpart receive reports and can therefore make a fresh decision of either X or Y for your second interaction. Knowing one another's initial choices, you each choose either X or Y and earn the payoffs indicated by Table 1.

### *Summary*

You will be asked to choose between actions XX or YY that will be taken in consecutive interactions with two different, anonymous, randomly chosen counterparts, each time generating earnings as shown in Table 1. After learning the outcome of your first interaction, you'll have the opportunity to report on your first counterpart's action at a cost to you of five cents. Second interactions proceed without fresh decisions, each individual's action being automatically the one initially chosen, unless the first interaction partner of either you or your new counterpart or both paid for reporting. If you receive a report about your new interaction partner's initial action and earnings, you'll be able to take a new decision and you'll also be told whether your new interaction partner received a report enabling him or her to take a new decision. If your new counterpart but not you receives a report, you will be informed of this and will simply wait while he or she makes their decision, after which the outcome will be reported to you.

### *Final details*

Once both interactions are completed and all participants have reviewed their final information screens, the main portion of the experiment will be over. At that point, you will be asked to answer some questions your answers to which will have no effect on your earnings. Remember that neither information about your decisions during the experiment nor your responses to these questions can be linked to you as an individual, since your decisions are recorded under a random identification number only. An experimenter will come to you with your payment in a closed envelope and you are then free to leave. No other participant will be told how much you earned in the experiment, and no participant will learn the identities of the two other participants with whom they were matched during the experiment.

Note that some further questions may appear on your screen while the decision portion of the experiment is in progress. In some cases, you may have an opportunity to add to your payoff by the accuracy of your answers.

## Comprehension questions

Before you make your decisions, we want to ask you some questions, which will appear in a moment on your computer screen. We hope that these questions will help you to check your understanding of the possible consequences of various choices. Please answer to the best of your ability. Your answers will not affect your earnings from the experiment, they will not prevent you from making any choice you wish to in the payoff-determining portion of the experiment, and they will have no impact on what participants the computer randomly assigns to interact with you.

(1) In the experiment, you are going to interact with two different participants in sequence. What is your first decision in this experiment?

(i) Choosing X or Y for interactions 1 and 2 from the set XX, XY, YX or YY, where the letter on the left (right) is the choice for the first (second) interaction. [            ]

(ii) Choosing X or Y for interactions 1 and 2 from among XX or YY only. [            ]

(iii) Making a choice of X or Y for the first interaction only. The choice for the second interaction comes later.

(2) Once every participant in the session has made their initial decision(s), you are given the opportunity to spend some amount of your earnings to reveal your first interaction partner's choice.

(2a) How much does it cost you to reveal your first interaction partner's choice? \_\_\_\_

(2b) Who in particular will find out your first interaction partner's choice if you decide to reveal it?

(i) Every participant in the session [            ]

(ii) Your first interaction partner's counterpart in the second interaction [            ]

(3) Suppose you reach the beginning of the second interaction. Suppose that you initially chose XX and that you are assigned to a counterpart for the second interaction who initially chose YY. Please answer the following questions.

a) Suppose that neither you nor your counterpart receives a report. What will you earn in the second interaction? \_\_\_\_\_ What will your counterpart earn in the second interaction? \_\_\_\_\_

b) Suppose, instead, that your new counterpart's first interaction partner chose to reveal information to you, but your own first interaction partner did not choose to reveal information to your new counterpart. You are free to change your choice of X or Y, but your counterpart is not free to change his or her choice of X or Y. What action do you think

that you would choose, knowing that your counterpart's choice is still Y? \_\_\_\_ What would you earn? \_\_\_\_ What would your counterpart earn? \_\_\_\_

(4) This question concerns initial decisions and their consequences for both interactions.

- a) What are your total earnings today (including \$5 for participation) if you and everyone else select XX and no reporting takes place? \$\_\_\_\_
- b) What are your total earnings today (including \$5 for participation) if you and everyone else select YY and no reporting takes place? \$\_\_\_\_.
- c) Suppose that all participants other than you choose XX. Given that that is the case, what should you choose if you want to earn as much as possible in the experiment as a whole and assuming that
  - (i) a participant who selects YY will be reported with high likelihood? \_\_\_\_\_
  - (ii) a participant who selects YY is very unlikely to be reported? \_\_\_\_\_
- d) Suppose that all participants other than you choose YY. Given that that is the case, what should you choose if you want to earn as much as possible in the experiment as a whole and
  - (i) a participant who selects XX will be reported with high likelihood? \_\_\_\_\_
  - (ii) a participant who selects XX is very unlikely to be reported? \_\_\_\_\_

**Table 1. Choices and Earnings from Each Interaction**

<b>If You Choose</b>	<b>&amp; Your Counterpart Chooses</b>	<b>You Earn</b>	<b>Your Counterpart Earns</b>
<b>X</b>	<b>X</b>	<b>\$10</b>	<b>\$10</b>
<b>Y</b>	<b>Y</b>	<b>\$5</b>	<b>\$5</b>
<b>X</b>	<b>Y</b>	<b>\$4</b>	<b>\$11</b>
<b>Y</b>	<b>X</b>	<b>\$11</b>	<b>\$4</b>